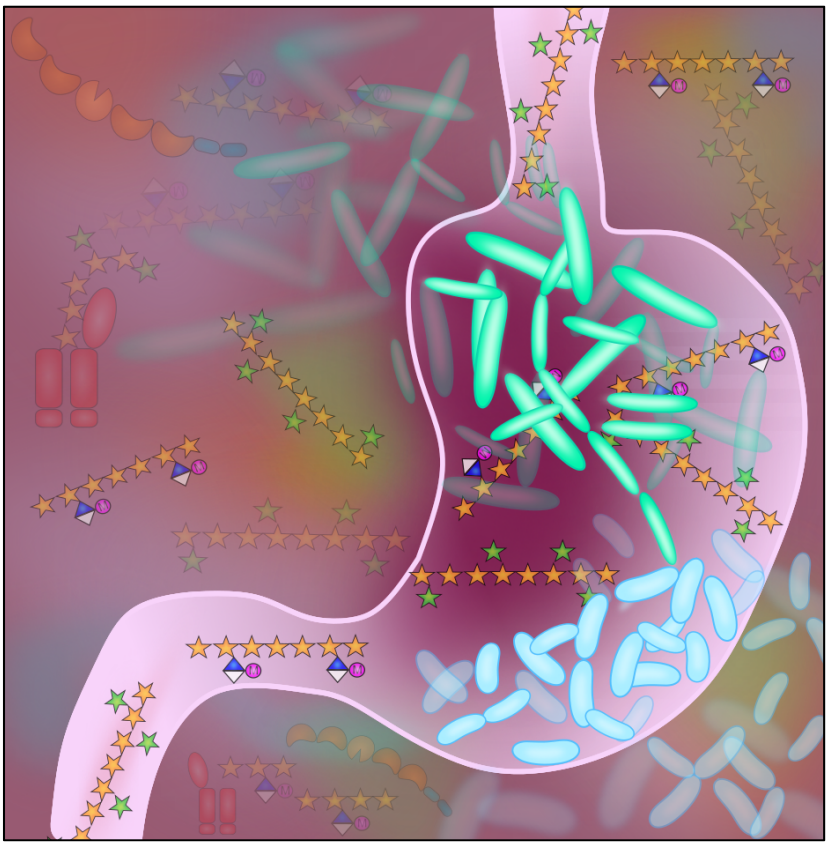
Glycan Utilization Strategy of the Butyrate Producing Gut Symbiont Roseburia intestinalis

Leth, Maria Louise

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Glycan Utilization Strategy of the Butyrate Producing Gut Symbiont *Roseburia intestinalis*



Maria Louise Leth

Ph.D. thesis

January 2019

DTU Bioengineering
Department of Biotechnology and Biomedicine

Glycan Utilization Strategy of the Butyrate Producing Gut Symbiont *Roseburia intestinalis*

Ph.D. thesis

Maria Louise Leth
Protein Glycoscience and Biotechnology
DTU Bioengineering

January, 2019

Kgs. Lyngby, Denmark. E-mail: malole@bio.dtu.dk

Preface

This Ph.D. thesis is the result of my work carried out from December 2014 to January 2019 under the supervision of Professor Maher Abou Hachem. The work has mainly been performed at Protein Glycoscience & Biotechnology, DTU Bioengineering at the Technical University of Denmark. Experiments have in addition been performed at Gut Microbiology and Immunology (DTU Food, Denmark), Infection Microbiology (DTU Bioengineering, Denmark), Department of Chemistry, Biotechnology and Food Science (NMBU, Norway) and NOBIPOL, Department of Biotechnology and Food Science, (NTNU, Norway). The project was supported by a Graduate School DTU Scholarship, Lyngby, Denmark, with additional support from Independent Research Fund Denmark, Natural Science (DFF, FNU) by Research Project 2 grant (Grant ID:4002-00297B) to MAH, a BIONÆR project(Grant ID: 244259) to BW and the Norwegian NMR Platform, NNP from the Research Council of Norway(Grant ID 226244) to FLA.

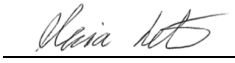
The project has resulted in the following publications and manuscript, which are included as chapters in the thesis.

- **Maria Louise Leth**, Morten Ejby, Christopher Workman, David Adrian Ewald, Signe Schultz Pedersen, Claus Sternberg, Martin Iain Bahl, Tine Rask Licht, Finn Lillelund Aachmann, Bjørge Westereng and Maher Abou Hachem. (2018). *Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut.* Nat. Microbiol. vol. 3, no. 5, pp. 570–580, 2018.

- Eva Madland, Yoshihito Kitaoku, Gerd Inger Sætrom, **Maria Louise Leth**, Morten Ejby, Maher Abou Hachem, Finn Lillelund Aachmann. (2018) *^{1}H, ^{13}C and ^{15}N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from Roseburia intestinalis*. Biomol. NMR Assign.

- **Maria Louise Leth**, Morten Ejby, Eva Madland, Yoshihito Kitaoku, Gerd Inger Sætrom, Dirk Jan Slotboom, Finn Lillelund Aachmann, Maher Abou Hachem (2019) *Structure and functional insight into a new xylan binding module of the multi-modular xylanase from Roseburia intestinalis.* (Manuscript in preparation)

- Sabina Leanti La Rosa, **Maria Louise Leth**, Leszek Michalak, Morten Ejby Hansen, Nicholas A. Pudlo, Robert Glowacki, Christopher Workman, Magnus Ø. Arntzen, Phillip B. Pope, Eric C. Martens, Maher Abou Hachem, Bjørge Westereng. *Molecular and Biochemical Basis for Complex β-Mannans Catabolism by the Human Gut Firmicutes Bacterium Roseburia intestinalis* (Accepted for publication in Nat. Commun. Jan. 2019)



Maria Louise Leth, Kgs. Lyngby, January 2019

Acknowledgements

During the years as a Ph.D. student I have received help and advice from a number of people, whom I would like to thank.

First and foremost I would like to thank **Professor Maher Abou Hachem** for great supervision and for giving me the opportunity to carry out my Ph.D. within an exciting and relevant research field. I appreciate the guidance, all the scientific ideas and our invaluable discussions.

I would like to thank **Morten Ejby Hansen** for his advice and guidance in the laboratory and for preforming some of the binding studies, and especially for his work on the crystallization of *Ri*CBMx and esterase *Ri*AXE2. I would further like to thank **Professor Dirk Jan Slotboom** and **Professor Leila Lo Leggio** for their collaboration on the crystallization structures. Leila is in addition thanked for her discussions on the *Ri*AXE2 structure, even though we did not succeed in solving the structure.

Laboratory technician **Marzanna Due** and Laboratory technician trainee **Thanh Holm Madsen** are sincerely thanked for their help in the cloning and initial binding studies of the transport protein *Ri*XBP.

Associate Professor Bjørge Westereng is thanked for a successful collaboration on the *Roseburia intestinalis* project. I am further thankful for his in assistance during my research stay in Ås, Norway and for his introduction to MALDI ToF.

Professor Finn Lillelund Aachmann and his group are thanked for the superb NMR analysis of *Ri*AXE and *Ri*CBMx and for writing the article "*^{1}H, ^{13}C and ^{15}N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from Roseburia intestinalis*".

I am very thankful for the effort **Associate Professor Christopher Workman** and **Ph.D. Adrian Ewald** have put into helping me with the analysis of the transcriptomic data.

Professor Tine Rask Licht is thanked for the collaboration on the completion studies at DTU Food and for good discussions. Many thanks to **Senior Researcher Martin Ian Bahl** for his help and guidance on the q-PCR work.

Associate Professor Claus Sternberg is thanks for the collaboration on fluorescence microscope analysis.

All my great present and former colleagues at Protein Glycoscience & Biotechnology and Enzyme, Protein & Chemistry are thanked for their help and scientific discussions. I will specifically like to thank my Cool Cat office mates **Mia, Morten, Susan** and **Jens** for their friendships and for making even the hard times fun. Thanks to **Karina** for being a great swimming buddie.

Thanks to **Michael** for his positive attitude and for helping with experiments during my pregnancy.

Finally, I would like to thank my family and friends for their endless support and understanding throughout difficult times. Thanks, to my son **Karl**. It has been fantastic to come home to a little guy, who is always happy and full of energy. A big thanks to **Mike** for the final proof reading of this thesis and for always believing in me. I couldn't have done it without your love and support.

Abstract

The human gut microbiota's (HGM's) interplay with the host and diet exerts a profound impact on our health. Dietary fibers, which are fermented to short chain fatty acids (SCFAs), are key in shaping the composition and the metabolic output of the HGM. The SCFA butyrate, produced mainly by *Clostridium* XIVa Firmicutes modulates host immune homeostasis and confers protection from inflammatory disorders and colorectal cancer. Despite this vital role on human health, the mechanisms underpinning the utilization of dietary fibers by butyrate producers are largely unexplored. This thesis aims at bringing insight into this facet using the abundant and prevalent butyrogenic bacterium *Roseburia intestinalis* as a model system.

The major dietary fiber xylan is shown to be an excellent substrate for *R. intestinalis* in this work. Transcriptomic analysis identified the molecular apparatus that confers growth on xylan, including a key cell-attached modular xylanase of glycoside hydrolase family 10 (GH10). This enzyme harbor a new low affinity xylan binding module, which showed an atypical architecture, featuring an open and shallow binding site with a few polar contacts to a single xylosyl unit. Investigation of transport and intracellular hydrolysis outlined a model for xylan breakdown to monosaccharides. An ATP-binding cassette (ABC) transporter with preference for 4-5 xylosyl units creates a competitive window that enables *R. intestinalis* to co-grow with key primary xylan degraders from the *Bacteroides* genus, which possess an uptake preference for larger ligands.

R. intestinalis uses a similar approach, as described above, to capture, degrade and transport complex dietary β-mannans. Capture and depolymerisation of mannan is initiated by a modular GH26 mannanase and the mannan-oligosaccharides are subsequently imported into the cell for further hydrolysis using an ABC uptake system. An *in vivo* experiment in mice colonized by a mock community of human commensals showed that β-mannan boosts commensal mannan degraders, including *R. intestinalis*.

Altogether, this project highlights the differentiation of capture and transport preferences of primary glycan degraders using xylan as a model. It provides a potential strategy for promotion of key members of a healthy HGM, by designing prebiotics that selectively target specific health promoting taxonomic groups.

Dansk resumé

Den menneskelige tarmmikrobiotas (HGMs) samspil med deres vært og kosten har stor indflydelse på vores helbred. Kostfibre, der bliver fermenteret til kortkædede fedtsyre (SCFA), er vigtige i forhold til kompositionen af denne population og dets metabolistiske produkter. Den kortkædede fedtsyre butyrat, som primært bliver produceret af *Clostridium* XIVa Firmicutes, modulerer menneskets immunhomøostase og giver beskyttelse mod inflammatoriske sygdomme og tarmkræft. På trods af den væsentlige rolle butyratproducenter har på vores helbred, er mekanismerne der understøtter nedbrydningen og optagelsen af kostfibre af denne gruppe af bakterier ikke særligt godt forstået. Denne afhandlings formål er at give indsigt i dette aspekt ved at bruge den udbredte butyratproducerende bakterie *Roseburia intestinalis* som et model system.

En transkriptionsanalyse identificerede det molekylære apparat, der er ansvarlig for vækst på xylan, herunder en stor celleforankret modulær xylanase fra glykosid hydrolase familie 10 (GH10). Dette enzym indeholder et nyt bindningsmodul med lav affinitet for xylan. Det viste sig at have en atypisk arkitektur med et åbent og fladt bindings site med få polære interaktioner til en enkel xylosylenhed. En undersøgelse af transportsystemet og intracellular hydrolyse enzymer skitserede en model for nedbrydningen af xylan til monosakkarider. En ATP-bindende kassette (ABC) transporter med præference for 4-5 xylosyl enheder kreere et vindue for at *R. intestinalis* kan gro sammen med den xylan nedbrydende *Bacteroides ovatus,* som har præference for større ligander.

R. intestinalis benytter sig af en lignende strategi til at fange, nedbryde og transportere kompleks β-mannans fra vores kost. Den indledende binding og depolymerisering af mannan er udført af en modulær mannanase, og herefter importres mannan-oligosakkariderne ind i cellen for videre nedbrydning via et ABC optagelsessystem. Et *in vivo* eksperiment, hvor mus blev koloniseret med et defineret samfund af gavnlige bakterier, viste at β-mannan kunne booste gavnlige mannannedbrydende bakterier, som *R. intestinalis.*

Ved at bruge xylan som model fremhæver dette studie differentiering af binding- og transportpræferencer hos glykannedbrydende bakterier. Dette giver en potentiel strategi for fremme væksten af vigtige medlemmer af en sund HGM ved at designe selektive prebiotika til specifikke taxa.

Abbreviations

ABC transporter	ATP-binding cassette transporter
AXOS	Arabinoxylan-oligosaccharides
AX	Arabinoxylan
CAZymes	Carbohydrate-Active enZymes
CBM	Carbohydrate binding module
CE	Carbohydrate esterase
CHX	Complex heteroxylan
DP	Degree of polymerization
GaGM	Galactoglucomannan
GaM	Galactomannan
GAX	Glucuronoarabinoxylan
GH	Glycoside hydrolase
GM	Glucomannan
GX	Glucuronoxylan
HDAC	Histone deacetylase
HGM	Human gut microbiota
HTCS	Hybrid two-component system
Interleukin-6	IL-6
Interleukin-12	IL-12
LM	Linear mannan
MFS	Major facilitator superfamily
MOS	Manno-oligosaccharides
NBD	Nucleotide-binding domain
PTS	Phosphotransferase system
PUL	Polysaccharide utilization loci
SCFA	Short chain fatty acid
SBP	Solute binding protein
SGBP	Surface glycan-binding protein
Sus	Starch utilization system
TBDT	TonB-dependent transporter
TMD	Transmembrane domain
XOS	Xylo-oligosaccharides

Contents

Aim of the present study

The work presented in this thesis focus on bringing molecular insight into the glycan capture, uptake and degradation machinery of the human gut symbiont *R. intestinalis*. This was done by identification and characterization of enzymes and transport proteins mediating glycan catabolism, with a focus on the hemicelluloses xylan and mannan as model substrates.

The scientific aims are:

- Establish growth profiles of *R. intestinalis* on the dietary fibers xylan and mannan.
- Identify potential gene loci encoding enzymes and transport proteins involved in xylan and mannan utilization using a transcriptomic based approach.
- Biochemical and structural characterization of key xylan utilization enzymes and transporters.
- Investigate the interplay between *R. intestinalis* and other primary xylan degraders from the human gut microbiota.

Chapter 1 Introduction

1.1 The human gut microbiota

The human gastrointestinal tract harbors a complex and dynamic community of microorganisms (collectively known as the human gut microbiota(HGM)), which plays an important role for its human host's health[1]. The HGM includes all three genetic branches of life; Bacteria, Archaea, Eukarya, with the former being the most abundant and diversely represented[2]. In a recent study, the number of bacteria in the HGM is estimated to be $3.8 \cdot 10^{13}$ bacteria - in a 70 kg "reference man", which equals a total mass of about 0.2 kg[3]. The study also updates the widely-cited 10:1 ratio, showing that the number of bacteria in the body is within the same order of human cells[3]. The adult HGM is dominated by members of two phyla; the Gram positive Firmicutes and the Gram negative Bacteroidetes. These phyla, together with the Actinobacteria, Proteobacteria, and Verrucomicobia, contain almost all of the bacterial species found in the human gastrointestinal tract[4]. Despite harboring bacterial genera from just a few phyla, the HGM shows a tremendous diversity at lower taxonomic levels. There are as such several hundred different species in the human gut, with at least 57 species detected in ≥90% of all individuals[5].

The composition of the HGM along the gastrointestinal tract is not homogenous. Physical variations including chemical and nutrient gradients have led to the development of specific microbial niches (Figure 1). In the gut, the small intestine has the lowest amount of bacteria and is dominated by a few species that tolerate the combined effects of bile acids and antimicrobials[2]. By contrast, the colon has the highest density of bacteria and a more diverse community, which makes it one of the most densely populated natural ecosystems. The intestinal epithelium is covered by a mucus layer of variable thickness and composition. In the colon, the mucosa is divided into an inner and an outer layer, with the latter providing a site of adhesion and a nutrient-rich habitat for adherent members of the HGM [1]. The wall of the colon contains folds, creating compartments between the folds (inter-fold region) that has higher amounts of mucus and are distinct from the central luminal (digesta) (Figure 1). In a mice study, the inter-fold regions were highly enriched in the phylum Firmicutes and more specifically, the families Lachnospiraceae and Ruminococcaceae[6]. Other families such as Prevotellaceae, Bacteroidaceae and Rikenellaceae were all enriched in the digesta region[6].

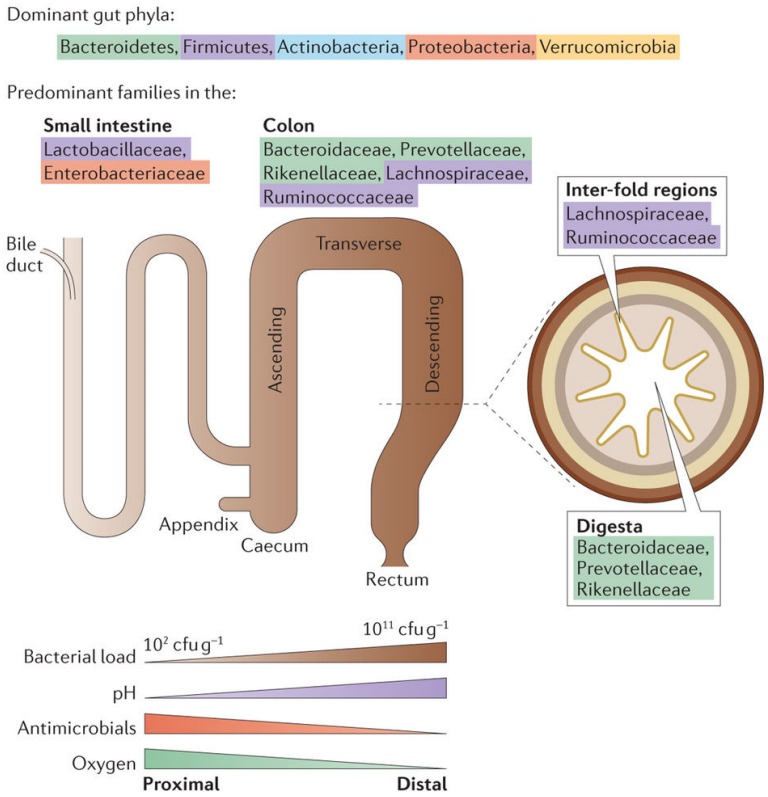


Figure 1 Microbial habitats in the human gut. In the small intestine, the families Lactobacillaceae and Enterobacteriaceae dominate, whereas the colon is characterized by the presence of species from the families Bacteroidaceae, Prevotellaceae, Rikenellaceae, Lachnospiraceae and Ruminococcaceae. Physiological differences exist along the length of the gut. For example, a gradient of oxygen, antimicrobial peptides (including bile acids, secreted by the bile duct) and pH limits the bacterial density in the small intestinal community, whereas the colon carries high bacterial loads. A cross-section of the colon shows the digesta, which is dominated by Bacteroidaceae, Prevotellaceae and Rikenellaceae, and the inter-fold regions of the lumen, which are dominated by Lachnospiraceae and Ruminococcaceae. Cfu, colony-forming units. Figure adapted from[2]

The neonatal gastrointestinal tract is believed to be sterile or contain a very few microbes[7]. The HGM is established shortly after birth and develops with the host through life. During the first year of life, the composition of the HGM is rather simple and varies to a larger degree among individuals[7]. The first colonizers of the gastrointestinal tract are facultative anaerobes that create a new environment, which further promotes the colonization of strict anaerobes such as *Bacteroides*, *Clostridia* and *Bifidobacterium* species[8]. The composition and diversity change to resemble that of an adult HGM by the end of the first 2 years of life with the dominance of Firmicutes and Bacteriodetes[8], [9]. Different factors can influence the composition of the adult HGM including the mother's HGM, mode of delivery (vaginal or caesarean), environmental exposition, antimicrobials, and diet[10]. The diet seems to be the major factor with the HGM responding rapidly to large changes in the diet. Clear differences have been observed in populations with a plant-based diet compared to a Western diet that is relatively low in

dietary fibers[11], [12]. A recent study showed that host genetics has a minor role in determining the composition of HGM and confirms the statement that environment and diet have the larger impact than host genetics[13].

The metagenome of the HGM is estimated to be approximately 150 times larger than the human genome[5] endowing the host with a substantial expansion of metabolic capabilities. The HGM can metabolize host indigestible carbohydrates, supply the host with essential nutrients and vitamins, and make minerals more accessible[14]. It also offers additional benefits to the host such as protection against pathogens[15], regulation of host immunity[16] and metabolic homeostasis[17]. Given this close symbiotic relationship it is not surprising that the HGM exerts a major impact on human health and physiology. Abnormalities in the composition of bacteria (dysbiosis) are therefore also associated with immune, metabolic and neurological disorders at both local intestinal and systemic levels. For example, inflammatory bowel disease is correlated with dysbiosis, characterized by decreased abundance of Clostridia and an overall reduction in bacterial diversity[18]. Childhood asthma is associated with low bacterial diversity during the first years of life[19]. The HGM associated with obesity is characterized by an increase in the Firmicutes to Bacteriodetes ratio and reduced diversity[20]–[22]. These examples indicate that the extent of bacterial diversity in the human gut might be an important factor for maintaining immune and metabolic homeostasis. This is consistent with studies showing decreased bacterial diversity in Western populations compared with those of hunter-gather populations, whom are not known to develop many of these diseases[23], [24]. The decline in diversity and loss of specific phylogenetic groups in the Western population might be a consequence of modern lifestyles, medical practices and processed food[11]. The disease examples also raise the question of whether dysbiosis is a cause or a consequence of diseases. Transplantation experiments, where the HGM from a diseased animal is grafted to germ-free healthy recipient, have shown that disease phenotypes can be transferred by the HGM. The experiments include colitis[25] and metabolic syndrome[26], both being diseases that are also known to be affected by host genetics and environmental factors[27]. A better understanding of the cause of dysbiosis is necessary to develop strategies to restore or boost a healthy bacterial community and homeostasis.

1.2 Short chain fatty acids and health effects of butyrate

The HGM produces a wide range of metabolites, including short chain fatty acids (SCFAs) that are utilized by the host and exert a range of health-promoting functions[28]. These SCFAs are the major products of anaerobic fermentation of carbohydrates and are necessary waste products, required to balance redox equivalent production in the anaerobic environment of the gut[29]. One health effect attributed to SCFAs is the reduction of the luminal pH that inhibits growth of pathogenic microorganisms and increases the absorption of nutrients[30]. The concentration of SCFAs varies longitudinally across the gut, with the highest levels in the cecum and proximal colon[31]. The most abundant SCFAs are acetate, propionate, and butyrate representing 90–95% of the SCFAs present in the colon, typically occurring in a ratio of 3:1:1[29], [30]. Other fermentation products such as lactate, succinate and 1,2-propanol are also produced, but do usually not accumulate in high concentrations, as they serve as substrates for other bacteria, including propionate and butyrate producers[28]. Acetate, propionate, and butyrate are rapidly absorbed from the gut lumen, but their subsequent distribution, fate and effects on host cell metabolism differ[32]. Acetate achieves high systemic concentrations in the blood and is produced by acetogentic bacteria and by most gut anaerobes, with the latter producing other reducing equivalents in addition to acetate[28]. Propionate and butyrate are produced by a distinct subset of bacteria, with propionate contributing to glycogenesis in the liver and butyrate being the preferred energy source for colonocytes. There are furthermore differences in how acetate, propionate, and butyrate interact with host proteins and receptors. The rest of the section below will focus on butyrate, as it is most often considered to have highest impact on human health[33], [34].

The ability to produce butyrate is widely distributed among Gram-positive bacteria in the colon[35]. The two main butyrate producing groups are the *Faecalibacterium prausnitzii*, belonging to the *Clostridium* cluster IV, and *Eubacterium rectale/Roseburia* spp., belonging to the *Clostridium* cluster XIVa[35]. It is produced from carbohydrates via glycolysis from the combination of two moles of acetyl-CoA to form acetoacetyl-CoA, followed by reduction to butyryl-CoA, which can form butyrate by two different pathways (Figure 2). One way is by butyryl-CoA:acetate CoA-transferase and the second is by phosphotransbutyrylase and butyrate kinase[36]. Only few butyrate producers

are known to use the butyrate kinase pathway, such as *Clostridiumbutyricum, Coprococcuseutactus,* and *Coprococcuscomes.* The butyryl-CoA:acetate CoA-transferase pathway requires consumption of external acetate, which can come from cross feeding on an acetate producing bacteria[37].

Butyrate's health effects are partly associated with its ability to inhibit the activity of histone deacetylase (HDAC) in colonocytes and immune cells. The inhibition promotes hyper acetylation of the histones, which has consequences for gene expression and cellular differentiation, including increase in total colonic regulatory T cells, downregulation of pro-inflammatory cytokines, such as interleukin-6 (IL-6) and interleukin-12 (IL-12) in colonic macrophages[32]. HDAC inhibition is also thought to promote apoptosis of colorectal cancer cells[32]. Butyrate appears to have a dual role, sometimes referred to as the "butyrate paradox" whereby it induces proliferation in healthy colonocytes but terminal differentiation and apoptosis in transformed cells. This can be explained by the fact that butyrate is the preferred energy source for colonocytes, while cancerous colonocytes prefer glucose. Compared to normal colonocytes that oxidize butyrate, butyrate is accumulated 3-fold in nuclear extracts from cancer cells, where it can act as an efficient HDAC inhibitor[38], [39].

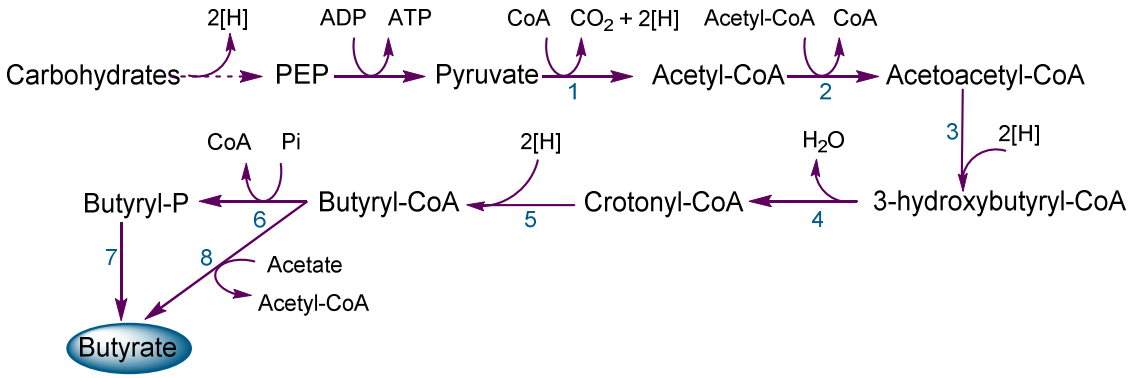


Figure 2 Pathway for microbial biosynthesis of butyrate from carbohydrates. Redox reactions which involve electron carriers are indicated by [H]. CoA, coenzyme A; P, bound phosphate; Pi, inorganic phosphate; PEP, phosphoenolpyruvate. Dotted line indicates that several intermediate steps are involved. Enzymes: 1, pyruvate-ferredoxin oxidoreductase; 2, acetyl-CoA-acetyltransferase; 3, β-hydroxybutyryl-CoA dehydrogenase; 4, 4-hydroxybutyryl-CoA dehydratase; 5, butyryl-CoA dehydrogenase; 6, phosphotransbutyrylase; 7, butyrate kinase; 8, butyryl-CoA:acetate-CoA transferase. Figure based on [28].

Several studies have shown that the number of butyrate producers are reduced in inflammatory diseases, irritable bowel syndrome[40], colitis ulcerosa[41], colon cancer[42], and Crohn's disease[43]. An explanation for the decrease of butyrate producers is not

yet clear, but the immune response might play a role in shaping the microbial community. Changes in the gut environment and transit time are in addition likely to have influence[35]. It is, however, becoming increasingly accepted that prominent butyrate producers are an indicator of a healthy HGM and seem to be actively involved in maintaining a stable and healthy community in the human gut[32].

1.3 *Roseburia intestinalis*

The genus *Roseburia* is one of the most abundant bacterial genera in the human gut and constitutes on average 5% of the total amount of the gut bacteria [35], [44]. In infants, the microbial abundance of *Roseburia* species is among the top 10 to increase within three month after birth[45]. *Roseburia* belongs to the Firmicutes phylum (Table 1), which is phylogenetically diverse. Analysis of this phylum, based on 16S rRNA gene sequences and multiple shared genotypic and phenotypic traits, has shown that the *Roseburia* genus cluster together with *Eubacterium rectale* forms a coherent group of butyrate-producing Firmicutes[35], [46]. The genus resides in the colon[37], [47] and has been shown to adhere to mucin, where it contributes to the mucosal butyrate production[48]. It is hypothesized that these bacteria play a critical role for human gut homeostasis and host epithelial health via the butyrate production in close proximity to the epithelium[48]. In recent years, an increasing body of evidence has shown a decrease in abundancy of *Roseburia* species in connection to chronic diseases such as type 2 diabetes[49], Chron's disease[50]–[52], and colorectal cancer[53]. This has led to the suggestion of using the abundance of *Roseburia* species as a health marker[54]. Colonization with *R. intestinalis* has in addition shown to decrease levels of inflammatory markers in atherosclerosis, as *R. intestinalis* produced butyrate, mediating these effects in a murine model[55].

Table 1 Classification of *R.intestinalis*

Domain	Bacteria
Phylum	Firmicutes
Class	Clostridia
Order	Clostridiales
Family	Lachnospiraceae
Genus	*Roseburia*
Species	*Roseburia intestinalis*
Strains	L1-82, M50/1, XB6B4

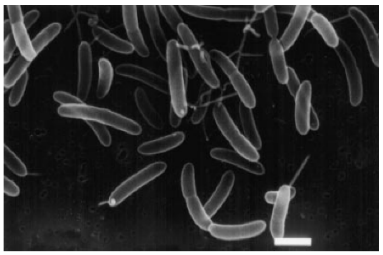


Figure 3. Scanning electron micrograph of *R. intestinalis* L1-82. Bar, 1 µm. Figure adapted from [57].

Five species of *Roseburia* have so far been described: *R. faecis, R. inulinivorans, R. hominis,* and *R. intestinalis*[56], [57]. This thesis focuses on the latter, which is one of the core 57 species that are present in ≥90% of all individuals[49]. *R. intestinalis* is a low-G+C-content Gram-positive to Gram-variable bacterium isolated from human feces[57]. It has a curved rod shaped morphology and measures approximately 0.5x1.5-5.0 µm with multiple subterminal flagella contributing to cell motility (Figure 3)[57]. *R. intestinalis* tolerance against oxygen has been tested, with exposure times above 2 minutes inhibiting growth, making it an obligate anaerobe bacterium[56]. Analysis of the fermentation products of *R. intestinalis* cells, grown on a rich media containing starch, glucose and cellobiose, showed that butyrate and lactate were the main SCFAs produced, with a concentration of 18.5 mM and 10.2 mM, respectively[57]. Formate was also detected but as a minor product in addition to hydrogen and carbon dioxide [57]. *R. intestinalis* can utilize acetate, however, the presence of this SCFA is not essential for growth. Interspecies transfer of acetate and hydrogen between microorganisms during xylan utilization has been demonstrated for *R. intestinalis*[58]. In this experiment, *R. intestinalis* was co-cultured together with an acetogen and a hydrogen-consuming bacteria (*Ruminococcus hydrogenotrophicus* and *Ruminococcus hydrogenotrophicus*) This led to a decrease in acetate and hydrogen and an increase of butyrate, indicating that *R. intestinalis* can utilize the acetate produced by the acetogen and that the hydrogen consumer can utilize the hydrogen from *R. intestinalis*[58]. The elimination of hydrogen is important, since it constitutes one of the main gases responsible for digestive discomfort for the host. However, hydrogen does not seem to affect the growth of *R. intestinalis*[59].

R. intestinalis ability to grow on a range of carbohydrate substrates has been tested (Table 2), In general, *R. intestinalis* is able to utilize a limited number of polysaccharides and monosaccharides compared to its counterparts from the *Bacteroides* genus[15]. A substrate that the strain grows particularly well on is xylan[56]. *R. intestinalis* has been described as a key xylan degrader in the human gut[60]. Xylanolytic bacteria were isolated from freshly voided feces obtained from 21 healthy individuals. The predominant xylanolytic bacteria (20 strains) were assigned to *R. intestinalis* and *Bacteriodes* species[60]. Compared to *Bacteriodes* species, *R. intestinalis* colonize insoluble xylan more efficiently. This was investigated by scanning electron microscopy, which showed that *R. intestinalis* colonize dietary fibers with a high density, forming a thick biofilm, while *Bacteriodes* cells are rather scattered on the dietary fibers[61]. The ability of *R. intestinalis* to attach to dietary fibers has been confirmed in another study with wheat bran

particles. The particles were incubated with fecal HGM from four individuals and washed to remove loosely attached bacteria. The results showed that the wheat bran particles were dominated by Lachnospiraceae, including *R. intestinalis*[62]. Recently, we have reported the molecular machinery of xylan utilization including enzymatic degradation and uptake preferences of *R. intestinalis*[63].

Table 2 Carbohydrates tested for utilization by *R. intestinalis*

Substrate	Growth		Substrate	Growth	
Arabinose	+	[57], [63], [64]	Amylopectin starch	+	[64]+ [65]
Fructose	+	[57]	Arabinan	-	[64]
Fucose	-	[64]	Arabinoxylan	+	[63], [64]
Galactose	+	[64], [66]	Cellobiose	+	[64]
Glucoronic acid	-	[63]	Fructooligosaccharides	+	[65]
Glucose	+	[63], [64]	Galactomannan	+	[63], [64]
N-acetylglucoamine	+	[64]	β-Glucan	-	[64]
N-acetyl-galactoamine	-	[64]	Glucomannan	+	[63], [64]
			Glucuronoxylan	+	[63]
			Glycogen	+	[64]
			Inulin	-	[57]
			Pectic galactan	-	[64]
			Pullulan	-	[64]
			Xyloglucan	-	[64]
			Xylo-oligosaccharides (DP 2-6)	+	[63], [65]

1.4 Fiber degradation by the microbiota

The human diet constitutes of large quantities of dietary carbohydrates and most of them cannot be exploited by the enzymes encoded by the human genome. The human digestive system can digest the diet-derived disaccharides sucrose and lactose, which are found in fruits and milk, respectively. The only dietary polysaccharide humans can digest is starch (mainly the α-1,4 bonds)[67]. Thus, most dietary carbohydrates reach the intestine intact and are termed dietary fibers[68](see next section for definition of dietary fibers). To access the energy from the otherwise indigestible dietary fibers, humans rely completely on the symbiotic HGM. The main products from this fermentation of dietary fibers by our HGM are SCFAs, which provide approximately 5-10% of the human energy requirements[69]. While the human genome only encodes 17 digestive enzymes, distinct gut bacteria produce hundreds of digestive enzymes targeting a variety of polysaccharides[68] (Figure 4).

The HGM composition and metabolic output are shaped by the influx of dietary fibers in the gut[70]. In a study, where humans over a short-term period of 4 days consumed

fiber rich or fiber poor diet compared to reference diet, it was shown that microbial community abruptly shift with a change in diet[71].

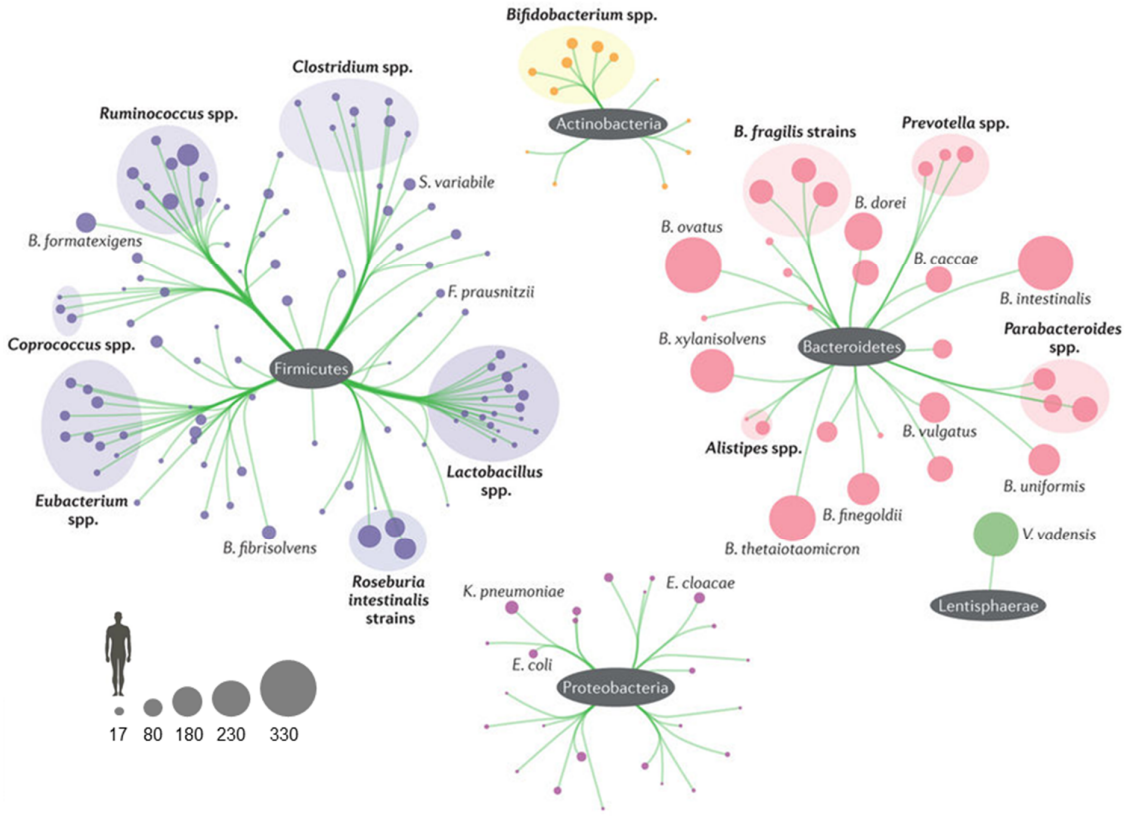


Figure 4: Diversity of dietary fiber active enzymes. Difference in the number of carbohydrate hydrolases (GHs) and polysaccharide lyases (PLs) between different taxonomic groups in a mini-microbiome and in a human is represented by the size of circle. The Bacteroidetes genomes in the mini-microbiome contain more GH and PL genes than genomes from the other phyla present, suggesting that members of this phylum are capable of using a larger range of carbohydrate substrates compared to the Firmicutes and the Actinobacteria. Figure modified from[68].

From an ecological perspective, the HGM can be divided into glycan utilization generalist or specialist; where generalist can utilize a broad range of dietary fibers by switching metabolism, whereas the specialist utilize only a few glycans[72]. Early studies on the two dominant phyla have shown that Bacteriodetes tend to be generalists and the Firmicutes specialists[73], [74], which also correlates with the number of carbohydrate active enzymes encoded in the representative genomes (Figure 4). Other factors that influence glycan fermentation by the HGM are transit time and the solubility of the dietary fibers[70] (Figure 5). In the ileum the transit time is fast and the bacteria target rapidly digestible dietary fibers, such as inulin and oligosaccharide side chains of pectin. In contrast, the transit time is slow in the distal colon and the dietary fibers are less soluble and therefore require longer time to digest. The density of bacteria is in addition

higher in the distal colon, which correlates with the slower transit time, where they have more time to proliferate.

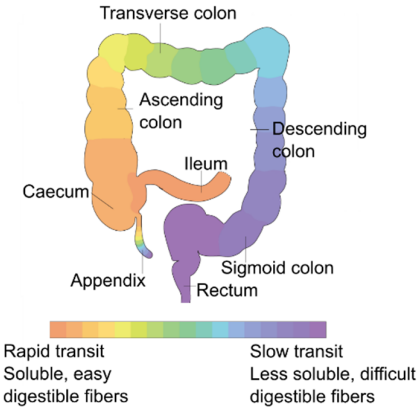


Figure 5: Dietary fiber utilization along the human gut. The schematic representation of the intestine is color coded to reflect the potential digestion gradients. Figure modified from[70].

1.5 Dietary fibers

The term dietary fiber has been defined by the 2009 Codex Alimentaroius as carbohydrate polymers with 10 or more monomeric units[1], which are neither digested nor absorbed in the small intestine[75]. Additionally, the definition states that added fibers must have been shown to have a physiological benefit to health[75]. Dietary fibers are primarily from plants, but can also originate from animals, fungal cell walls and milk-oligosaccharides. This section will focus on dietary fibers from plants, because cereals, fruits and vegetables are major components of human diet. Dietary fibers from plants include the fractions of starch that are inaccessible to human α-amylases (resistant starch) together with non-starch polysaccharides; gums, mucilages, inulin, and components of the plant cell wall[76]. The plant cell wall consists of cellulose microfibrils and pectins that are embedded in a matrix of hemicellulose (Figure 6). The structural features of these dietary fibers are diverse with respect to monosaccharide composition, anomeric configuration, glycosidic linkage-type, linear chain length, branched chain composition, reducing terminal attachments and modifications such as acetylations[77]. Even though, dietary fibers in general are composed of a relatively small number of monosaccharides (mainly glucose, xylose, arabinose, mannose, galactose, fructose, rhamnose, fucose, and the uronic form of some of the monosaccharides), the possible linkages

[1] Decision on whether to include carbohydrates with a degree of polymerization from DP 3 to 9 are left to national authorities. European Union has decided to include DP 3-9.

between these units are numerous, making dietary fibers some of the most chemically diverse polymers in nature. Thus, degradation of these will require several linkage-specific enzymes. Moreover, the polymers can be assembled in a way where one type is not accessible without prior degradation of another[67].

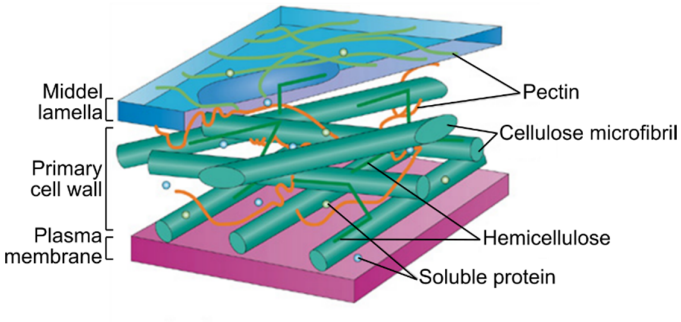


Figure 6: Plant cell wall structure. Diagrammatic representation of the major structural components; cellulose, hemicellulose, pectin, lignin and soluble proteins, of a "typical" plant cell wall. Figure modified from[78].

1.6 Hemicellulose

Hemicellulose are plant cell wall derived dietary fibers that reach the intestine intact. However, the digestibility of hemicellulose by the HGM is high. In a study with seven women on a standardized diet, it was estimated that 72% of the digested hemicellulose was degraded by the HGM[79].

In the plant cell wall hemicellulose constitute approximately one-third of the dry mass and contribute to mechanical strength by interactions with cellulose and, in some walls, with lignin[80], [81]. Hemicellulose are a heterogeneous group of polysaccharides that can be divided into the major groups; xylans, mannans, xyloglucans, and β-glucans[81]. These polysaccharides are characterized by having β-(1,4)-linked backbones of glucose, xylose, or mannose in an equatorial configuration, with the exception of β-glucan that contains both β-(1,4) and β-(1,3) linkages. The backbone is often decorated with sugars, which can be simple (a single monosaccharide, few linkage-types) or very complex (many monosaccharides, many linkage-types, and varying length of branches). Hemicellulosic polysaccharides are in addition acetylated, explaining why some of these polymers are not crystalline[82]. The fine structure of these polysaccharides depends on the decoration and varies in different plant tissues and species. Table 3 provides an overview of the backbone and decorations found in the different hemicellulose polysaccharides. The detailed structures of xylans and mannans will be discussed in the following section.

Table 3: Chemical structure of polysaccharides from hemicellulose

Polysaccharide	Backbone	Decoration
Xylan	β-(1,4)-linked D-xylose	Linear or branched with L-arabinose and/or D-glucuronic acid. D-xylose, D/L-galactose and ferulic acid or *p*-coumaric acid can also occur as esterification of arabinosyl residues.
Mannan	β-(1,4)-linked D-mannose	Linear or branched with D-galactose
Glucomannan	β-(1,4)-linked D-mannose and D-glucose	Linear or branched with D-galactose. Esterification of glycosyl residues can also occur.
Xyloglucan	β-(1,4)-linked D-glucose	Branched with D-xylose. D-galactose and L-fucose can also occur.
β-glucans	β-(1,4)- and β-(1,3)-linked D-glucose	Linear

1.7 Xylan

Xylan is the major component of hemicellulose and second most abundant polysaccharide in plant cells. It is found in high amount in hardwoods, where xylans can comprise 25-35% of the biomass, as well as in cereals grains which contain 30-50%[83]. It is typically found in the secondary cell wall of plants, but is also present the primary cell wall—particular in monocots[84]. Xylan shows large heterogeneity and it varies greatly in structure depending on plant origin. A common feature in xylans are a backbone of β-1,4-linked D-xylosyl residues with different side chain substitutions, such as D-glucuronic acid, 4-*O*-methyl-D-glucuronic acid, L-arabinose, D/L-galactose, D-xylose [81], [85]. Xylans can be grouped according to their side chains (Table 4) and its degree of polymerization is variable, such as in hardwood and softwood, with 150-200 and 70-130 β-1,4-linked D-xylosyl residues, respectively[84].

Glucuronoxylan (GX) is found in the secondary cell wall of dicots[81] and is characterized by having side chains of D-glucuronic acid and/or 4-*O*-methyl-D-glucuronic acid attached at position 2 of the D-xylosyl polymer. In hardwood and herbs the molar ratio of D-xylose and 4-*O*-methyl-D-glucuronic acid is on average 1:10. D-glucuronic acid substitutions are less common but has been reported isolated from husk of red gram[86]. The backbone of GX is usually acetylated at C2 and/or C3, with the latter being more frequent. The degree of acetylation varies considerably, e.g. birchwood GX is heavily acetylated with more than 1 mol of acetic acid per 2 mol of D-xylose[87]. The presence of acetyl groups is responsible for the partial solubility of xylan in water. GX has a unique sequence at the reducing end, which includes D-rhamnose and D-galacturonic acid (Figure 7) and it has been shown to be required for normal GX synthesis in the

secondary wall of dicots[88]. It is to this date unknown if other types of xylans have a similar sequence at their reducing ends.

Table 4: Structure and occurrence of xylan in plants. Modified from[89]

Polysaccharide	Backbone substitutions	Occurrence
Arabinoxylan (AX)	L-arabinose. Ferulic acid or *p*-coumaric acid can be *O*-esterified to L-arabinose residues.	Seeds of grasses and cereal grains
Glucuronoxylan (GX)	D-glucuronic acid that can be 4-*O*-methylated	Secondary walls of dicots; hardwoods, wood plants and herbs (methylated form). Various dicot seeds and fibers (methylated form).
Glucuronoarabinoxylan (GAX)	L-arabinose and D-glucuronic acid that can be 4-*O*-methylated. Ferulic acid or *p*-coumaric acid can be *O*-esterified to L-arabinose residues. Complex GAX is more heavily substituted and also includes D-xylose and D/L-galactose.	Minor component of dicot primary cell wall, hemicellulose in grasses, softwood. Complex GAX is mainly found in seeds, corn bran, gum exudates, and mucilages

Arabinoxylan (AX) is a dominant hemicellulose component of cereal grains and seeds from other monocots, such as rye grass, bamboo shoots, and pangola grass[90]. The AX D-xylose backbone contains L-arabinosyl side chains substituted to C2 and/or C3. This di-substitution allows high rates of L-arabinosyl substitutions, which in rye grain can reach up to 1.3 L-arabinosyl per D-xylosyl[91]. The backbone of AX can be acetylated at C2 and/or C3[92]. Another structural feature, is the presence of ferulic acid and *p*-coumaric acid, esterified on the C5 position of L-arabinosyl residues, which is found to crosslink AXs. This crosslinking is hypothesized to contribute to resistance to enzymatic degradation of plant cell walls[93].

Xylans can also possess both L-arabinosyl and 4-*O*-methyl-D-glucuronic acid substitution and are referred to as glucuronoarabinoxylan (GAX), which is found in lignified tissues of grasses and in softwoods[83]. In softwood the GAX is not acetylated and the 4-*O*-methyl-D-glucuronic acid content is higher compared to GX from hardwoods[87]. The structurally most diverse xylan is complex GAX that is present in seeds and corn bran. The backbone is highly substituted with mono-, di-, and trisaccharide side chains of D-xylosyl, 4-*O*-methyl-D-glucuronic acid, L-arabinosyl, and D/L-galactosyl [83], [89]. The backbone of complex GAX can be acetylated at C2 and/or C3. The distinction

between AX, GAX, and complex GAX is in the literature not clear and complex GAX is often called AX, GAX, or complex heteroxylan[85], [89], [94].

Most xylans occur as heteropolysaccharides with different substituent groups on the backbone—as with the tree major classes of xylan described above. Homoxylans, on the other hand, are linear and consist exclusively of D-xylosyls linked by 1,3- and/or 1,4-linkages and are common in algae and seaweed, but are rarely found in higher plants[83].

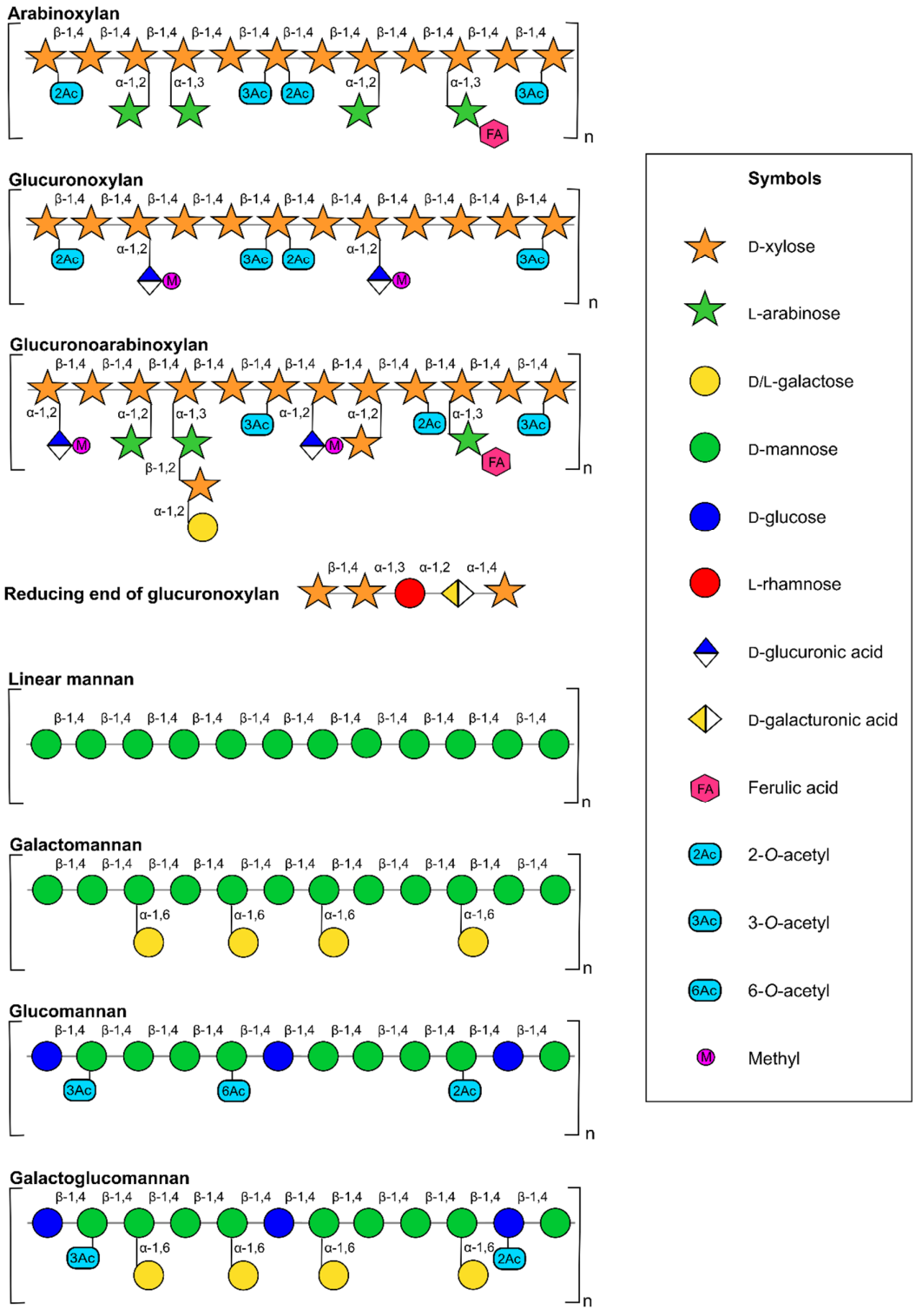


Figure 7: Structures of xylans and mannans. The monosaccharides and linkages in the main classes of xylan and mannan are represented.

1.8 Mannan

Mannan is the main hemicellulose in the primary cell wall of softwoods and in specialized structures in seeds and fruits[95]. The physiological roles of mannans are either structural as a component of the hemicellulose-cellulose network or as a storage polysaccharide in seeds[89]. The mannan backbone consist of β-1,4-linked D-mannosyl residues that can be interrupted by D-glucosyl and/or branched with D-galactosyl. Based on the backbone composition and the side chain substitution mannans can be grouped into four categories (Table 5).

Table 5 Structure and occurrence of mannans in plants.

Polysaccharide	Backbone and sidechain substitutions	Occurrence
Linear mannan (M)	Backbone of D-mannose. No sidechain substitutions	Marine algae, seed of e.g. ivory nut and coffee bean
Galactomannan (GaM)	Backbone of D-mannose branched with single D-galactose	Cell wall of e.g. legume seed (guar, carob, cassia)
Glucomannan (GM)	Backbone of D-mannose and D-glucose. No sidechain substitutions	Softwoods, hardwoods, grasses, herbal plants
Galactoglucomannan (GaGM)	Backbone of D-mannose and D-glucose branched with single or double D-galactose	Softwoods, hardwoods, grasses, herbal plants

Linear mannan (LM) is an important structural component of some marine algae and in seeds of many plants such as ivory nut and coffee bean[96]. It is a homopolysaccharide composed of β-1,4-linked D-mannosyl with no sidechains making it water insoluble.

Galactomannan (GaM) is main storage carbohydrate in the cell wall of seeds[90]. As in LM the backbone is composed of β-1,4-linked D-mannosyl, but is branched to various degree at C6 with single D-galactosyl residues. True GaM are mannans with more than 5% D-galactosyl and some branched types can have as high a mannosyl:galactosyl ratio as to 1.1:1[96], [97]. An increased D-galactosyl substitution makes GaM water-soluble and the most heavily substituted GaM are commercially used as gums (guar, tara gum, and carob).

Glucomannan (GM) is the main component of hemicellulose in the secondary cell wall of softwood and a minor component of hemicellulose of hardwood, grasses and herbal plants (3-5% of the wood dry weight)[95]. The backbone of GM contains randomly

arranged β-1,4-linked D-mannosyl and D-glucosyl and the ratio of the two sugars in the backbone varies depending on plant origin. GM can be acetylated at position C2 and/or C3, but acetylation at position C6 have also been reported[97]

GM branched with single D-galactosyl residues is termed galactoglucomannan (GaGM). The D-galactosyl residues can be attached to both D-mannosyl and D-glucosyl. There are two types of GaGM; galactosyl rich and galactosyl poor, where the molecular ratio of galactosyl:glucosyl:mannosyl is approximately 1:1:3 and 0.1:1:3, respectively[97]. Like GM, GaGM is water-soluble due to its relatively large content of D-galactosyl.

1.9 Carbohydrate active enzymes

Glycans and glycoconjugates are widely distributed in nature, where they mediate a variety of biological functions, from carbon reserves, to structural molecules, decorations on lipids and as a variety of specialized metabolites that mediate cellular recognition, communication and bioactivities. These functions are facilitated by a diverse chemical composition based on a large number of combinations of monosaccharides structures and different intersugar linkages, making them some of the most structurally diverse substrates on earth. This diversity is a product of multiple enzymes involved in the assembly, breakdown and/or modification of glycans. Collectively these enzymes are designated Carbohydrate-Active enZymes (CAZymes), which are classified into families in the CAZy database (www.cazy.org)[98]. The classification is based on sequence similarities and members of the same family display a common fold with a conserved catalytic apparatus and mechanism[82], [99]. To date, the CAZy database is divided into five enzymatic classes (glycoside hydrolases (GHs), polysaccharide lyases, carbohydrate esterases (CEs), and auxiliary activities) and one non-catalytic family (carbohydrate binding modules (CBMs). CAZymes are frequently modular enzymes with a catalytic module harboring a variable number of other modules, which can be either catalytic or not[98]. Thus, a modular CAZyme can be assigned to several families if its constitutive modules belong to separate families. In the following sections the CAZymes families involved in this thesis will be thoroughly described.

1.10 Glycoside hydrolases

Glycoside hydrolases (GHs) are widespread in nature and catalyze the cleavage of glycosidic bonds between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety[100]. They are important for biotechnological and biomedical applications and are so far the best characterized CAZymes. This is reflected by GHs constituting the largest category in the CAZy database with 154 families (January 2019). A number of the enzyme families are further categorized based on conservation of their tertiary structure, catalytic residues, and catalytic mechanism forming clans within the family, into groups that are named from A-N[101]. GHs are also frequently classified as "exo" or "endo" acting, which refers to the enzymes ability to cleave a polysaccharide from the end (often the non-reducing end) or within the polymer chain (Figure 8)[100].

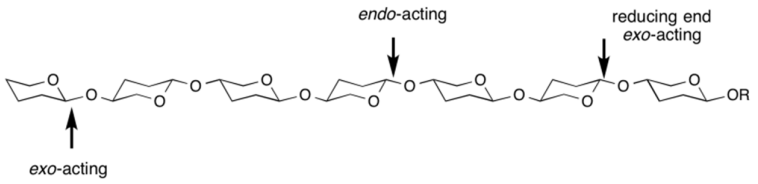


Figure 8 Schematic representation of exo- and endo-acting enzymatic cleavages of a polysaccharide. Figure adapted from https://www.cazypedia.org/index.php/Glycoside_hydrolases

GHs typically catalyze the hydrolysis of a glycosidic bonds via two general mechanisms resulting in either in overall inversion or retention of the anomeric configuration at the hydrolysis site (Figure 9)[102].

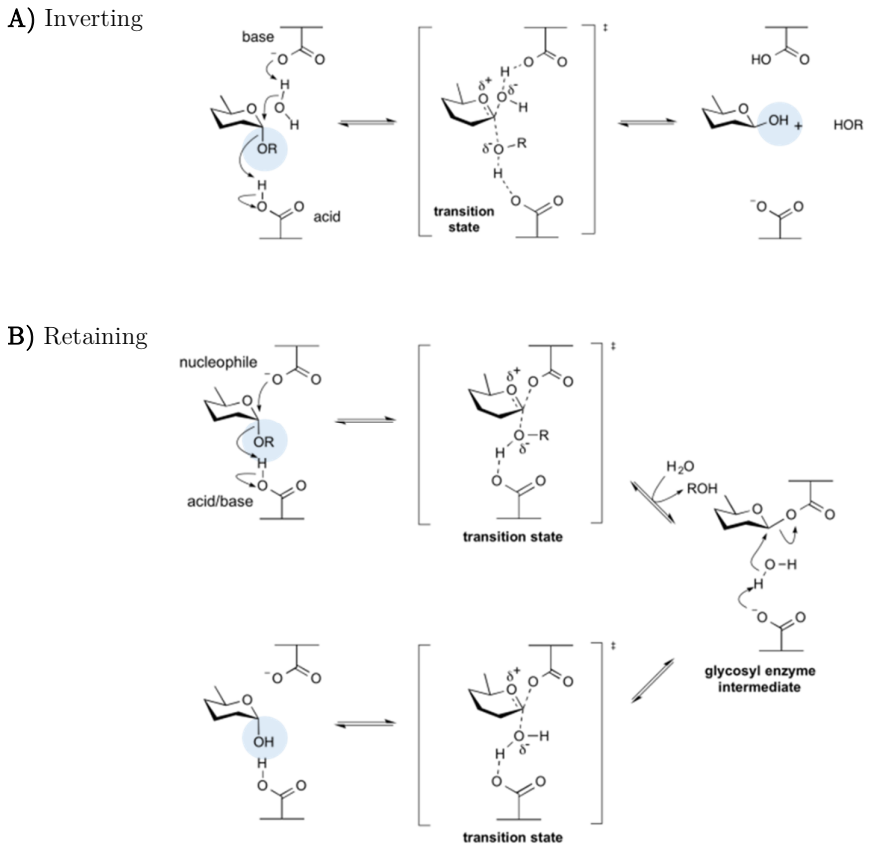


Figure 9: The two major mechanisms of enzymatic glycosidic bond hydrolysis. A) Inverting mechanism of α-glycosidase. B) Retaining mechanism of α-glycosidase. Figure modified from https://www.cazypedia.org/index.php/Glycoside_hydrolases

1.11 Carbohydrate binding modules

Many CAZymes are modular proteins composed of a catalytic module joint to one or more ancillary modules often by means of a flexible linker. The most common of ancillary modules are carbohydrate binding modules (CBMs). CBMs are defined as a contiguous, independently folding domains that are appended to catalytic CAZymes to confer binding to carbohydrate substrates[103]. The most recognized functions of CBMs is to mediate increased substrate binding, thereby bringing the appended catalytic module into close proximity and providing prolonged interaction with the target substrate[99], [104]. CBMs may contribute to catalysis by introducing conformational changes to carbohydrates[105]. An early report ascribed a disruptive function of crystalline cellulose to certain CBMs[106], however, evidence for this hypothesis is limited[107]. There are currently 84 CBM families in the CAZy database (January 2019),

with the majority targeting components of the plant cell wall. Structures for 63 of the CBM families have been determined, with most of CBM's assuming one of the seven "fold families" - with the β-sandwich being the most common fold[104]. To provide additional functional relevance to the CBM classification the modules have been grouped into three types; A, B, and C (Figure 10), which where updated in 2013[107]. Type A binds crystalline surfaces, type B binds internally onto single glycan chains (endo-type) and type C binds the terminal part of glycans (exo-type) (Figure 7).

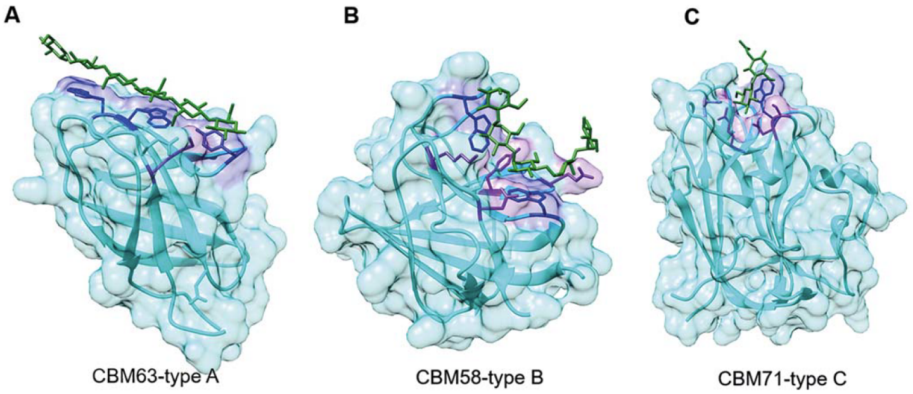


Figure 10: CBM types. A) CBM63 (PDB 4FER), B) CBM58 (PDB 3K8L), C) CBM71 (PDB 4CUB). The backbone is displayed as a cyan cartoon with a transparent solvent-accessible surface. Aromatic amino acids establishing CH– interactions are shown in blue color, amino acids forming hydrogen bonds are shown in purple (with the side chains displayed in sticks), and the substrate is shown in green. Figure adapted from [105].

Substrate binding are primarily facilitated by both the aromatic amino acids side chains residues and loop structures that shape the binding site to mirror the conformation of the ligand[104]. The aromatic residues are frequently tryptophan, tyrosine and less commonly, phenylalanine. These aromatic residues mediate binding to the substrate through hydrophobic and CH–π interactions, and the hydrophobic platforms at the binding site can be planar, have an angle that reflects the symmetry of the substrate (e.g. helical 3-fold symmetry for xylan) or form a sandwich[104], [105]. Because the binding site mirrors the substrate, the main feature of type A CBMs is planar with a hydrophobic surface capable of binding insoluble substrate, such as crystalline cellulose and chitin[108]. Type A CBMs have only been reported to have little or no affinity for oligosaccharides[59], [109] and the substrate binding is driven by a favorable entropy change consistent with a lower density of hydrogen bonds per sugar ring. Type B CBMs (endo-type), in contrast, possess a binding site in shape of a cleft that accommodates the substrate[105], [107]. The binding site can be located in either the variable loop site, positioned towards the end of the β-sandwich, or in the concave face site, with the β-

strands perpendicular to the ligand chain. The latter region facilitates binding in most cases, however, there are families with two binding sites that involve both regions. In general, type B CBMs binds a variety of glycans, such as xylans, mannans, starch and galactans[108], [110], [111] and the binding is dominated by a favorable enthalpy change. Type C CBMs (exo-type) display a binding site with pocket topology consistent with the recognition of the non-reducing end of glycans[108]. Most frequently type C CBMs recognize exposed or short glycans, such as monosaccharides, disaccharides and disaccharides[105].

The substrate specificity for a given CBM family can be towards a single type of glycan (CBM17, CBM27) or it can be diverse (CBM6, CBM35, CBM4). The diversity in substrate recognition contributes to targeting efficiencies of an enzyme in environments with a range of glycans present, such as the cell wall[82]. Often, but not invariably, CBMs display a binding specificity that reflects the activity of the catalytic module. In general, CBM-glycan interactions are quite weak (K_D affinities in μM^{-1} to mM^{-1} range) indicating that very tight binding is not necessarily an advantageous for enzymatic activity. However, these weak interactions are frequently compensated by the presence of several and/or multiple copies of CBM associated to a given enzyme, which can be a way of fine tuning the specificity to the substrate or the dynamics of enzyme binding (affinity/turnover ratio). Thus, diversity in substrate specificity, tandem or multiple CBMs are together with binding site topology important factors in understanding the mechanisms of substrate recognition of a CBM containing enzyme.

1.12 Carbohydrate esterases

A large group of enzymes are CEs that catalyze the *O*- or *N*-deacylation of ester based modifications present in mono-, oligo- and polysaccharides[99]. The CAZy database has presently classified CEs into 16 families (CE1-CE16), however, there are currently no members of family CE10. The substrate specificity of CEs is very diverse (Table 6) and as the barrier between carbohydrate esterases and other esterase activities is low, it is likely that the sequence-based classification incorporates some enzymes that may act on non-carbohydrate esters[99].

Table 6 Overview of carbohydrate esterase families. Modified from[112]

CE family	Catalytic residues	Known activities	Structural fold
CE1	S,H,D	Acetylxylan esterase, feruloyl esterase, carboxylesterase, S-formylglutathione hydrolase, diacylglycerol *O*-acyltransferase, and thehalose 6-*O*-mycolyltransferase	α/β hydrolase fold
CE2	S,H	Acetylxylan esterases	Repeated α/β/α motifs; C-terminal SGNH domain and an N-terminal jellyroll domain
CE3	S,H,D	Acetylxylan esterases	Repeated α/β/α motifs; N-terminal SGNH domain and linker module on C-terminal
CE4	D,H,H,Co^{2+}/ Zn^{2+}	Acetylxylan esterases, chitin deacetylase, chitooligosaccharide deacetylase, peptidoglycan *N*-acetylglucosamine deacetylase, and peptidoglycan *N*-acetylmuramic acid deacetylase	Distorted $(\alpha/\beta)_7$ barrel fold; NodB domain
CE5	S,H,D	Acetylxylan esterases and cutinases	α/β hydrolase fold
CE6	S,H,E	Acetylxylan esterases	Repeated α/β/α motifs; SGNH domain
CE7	S,H,D	Acetylxylan esterase and cephalosporin-C deacetylase	α/β hydrolase fold; hexameric
CE8	D,DR	Pectin methylesterase	Right-handed -helix fold
CE9	D, Fe^{2+}/ Zn^{2+}/ Co^{2+}/Cu^{2+}	*N*-acetylglucosamine-6-phosphate deacetylase	$(\alpha/\beta)_8$ barrel
CE11	E, Zn^{2+}, Zn^{2+}	UDP-3-*O*-acyl-*N*-acetylglucosamine deacetylase	Two-layer-sandwich
CE12	S,H,D	Pectin acetylesterase, rhamnogalacturonan acetylesterase and acetylxylan esterase	Repeated α/β/α motifs; SGNH domain
CE13	S,H,D	Pectin acetylesterase	Unknown
CE14	H,D, Zn^{2+}	*N*-acetyl-1-D-*myo*-inosityl-2-amino-2- deoxy-D-glucopyranoside deacetylase (MshB, mycothiol S-conjugate amidase and *N,N'*-diacetylchitobiose deacetylases	α/β fold; Rossmann fold
CE15	S,H,E	Glucuronoyl esterase	α/β fold; Rossmann fold
CE16	-	Acetylesterase	Unknown

1.13 Enzymatic degradation of xylan

The structure of xylan is heterogeneous and complex (section 1.7). Thus, the complete breakdown of xylan requires several enzymes with diverse specificity and mode of action. Xylanolytic enzymes can be classified into two main groups: those acting on the xylose backbone, *i.e.* endo-β-1,4-xylanases and β-D-1,4-xylosidases, and those cleaving the side

chains, *i.e.* α-L-arabinofuranosidases, α-1,2-glucuronidases, acetylxylan esterases, feruloyl esterases, *p*-coumaroyl esterase, β-D-1,2-xylosidases, and α-D/L-galactosidases (Figure 12 and Table 7). All of these enzymes act cooperatively to degrade xylan into its constituent sugars[87].

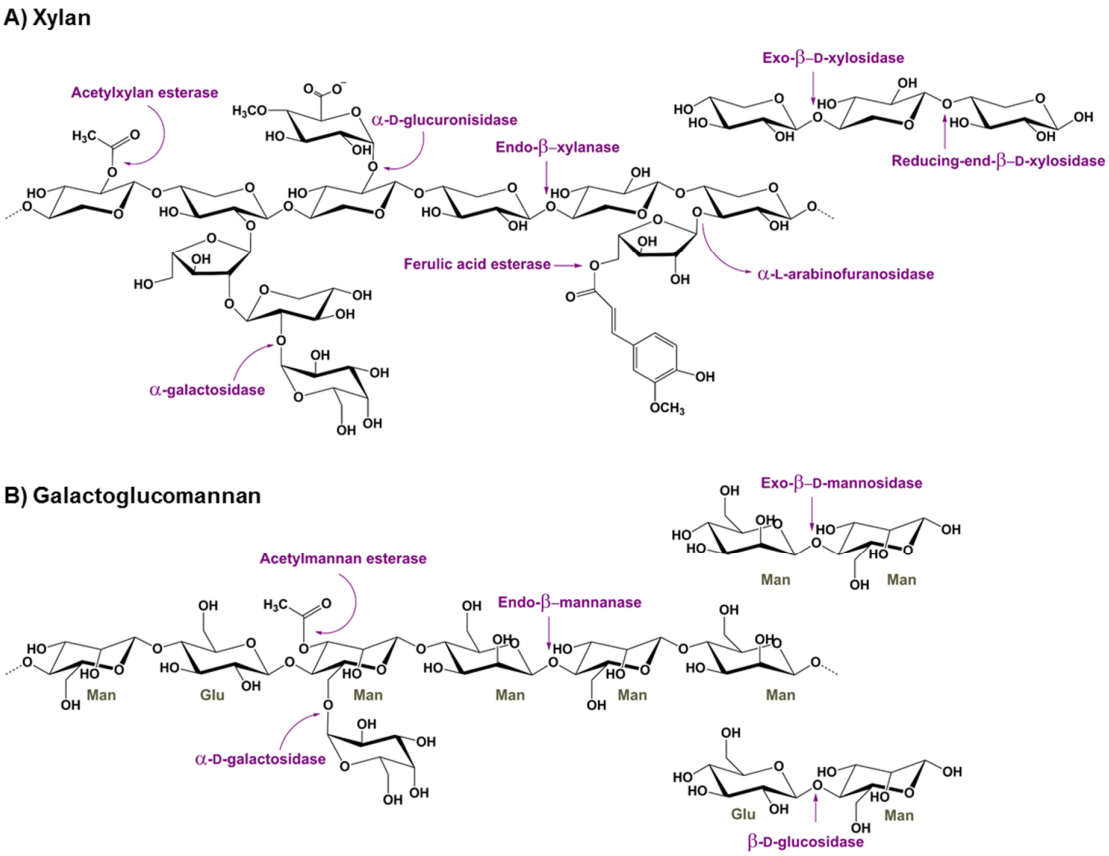


Figure 12. Xylanolytic and mannolytic degrading enzymes. The site of enzymatic attack on xylan and mannan is indicted by an arrow.

An important enzymatic activity for depolymerization of xylan is endo-β-1,4-xylanase activity. These enzymes cleave the β-1,4 glycosidic linkage between xylosyl residues in the backbone of xylan[113] and are primarily archived by xylanases classified into GH families 10, 11, and 30[82], [114]. The cleavage mode of these families are different and differs mainly in the way the recognition site forms the enzyme-substrate complexes. GH10 requires two unsubstituted xylosyl residues at subsites[2] -1 and -2 to attack the

[2] Subsites are the sites at which enzyme interacts with the substrate (sugar) in the active site. The cleavage of the glucoside bond takes place between subsite -1 and +1. Subsites are numbered with increasingly negative numbers (-1, -2, -3, etc.) away from the cleavage point towards the non-reducing terminus, and with increasingly positive numbers (+1, +2, +3, etc.).

xylan backbone and can cleave xylosidic linkage to a decorated xylosyl residue at subsite +1[114]. Enzymes of GH11 require three unsubstituted xylosyl residues at subsite -1, -2 and -3 and can cleave the backbone xylosidic linkage with a decorated xylosyl at subsite +2[114]. Thus, the product profiles from hydrolysis with these two GHs will be different - GH10 hydrolysis yields xylo-oligosaccharides (XOS) with terminal non-reducing end decorations, while GH11 yields XOS with decoration at the penultimate xylosyl from the non-reducing end[113]. GH30 xylanases are specialized for GX and their mode of action is determined by the presence of glucuronic acid or 4-*O*-methylglucuronic-acid. In general GH30 attack the GX backbone xylosidic linkage with a decorated 4-*O*-methylglucuronic acid xylosyl residue at subsite -2, which results in hydrolysis products that are XOS with 4-*O*-methylglucuronic acid decoration at the penultimate xylosyl residue to the reducing end[114], [115].

Other xylan backbone acting enzymes are exo-β-1,4-xylosidases that release xylose monomers from the non-reducing end of XOS. The affinity of exo-β-1,4-xylosidases for XOS decreases with increasing DP[116]. Exo-β-1,4-xylosidases are classified into five families; GH3, 30, 39, 43 52 and 120, which are retaining expect for GH43 (inverting family). Another type of xylosidases are reducing-end-β-1,4-xylosidases classified as GH8 that release xylose from the reducing end of XOS and are active on linear and decorated XOS [117], [118]. Besides the accommodation of decoration in GH8, another difference between these families of xylosidases is that the reducing-end-β-1,4-xylosidase cannot hydrolyze XOS< xylotriose[63].

The α-1,2 linkage of glucuronic acid and 4-*O*-methylglucuronic acid residues to the backbone is one of the most acid stable glycosidic linkages in the plant cell and persist in an acid hydrolysate of GX[114]. However, the residues are recognized and can be liberated from the backbone by α-1,2-glucuronidases. There is two classes of α-1,2-glucuronidases; GH67 and GH115. Members of GH67 can only release 4-*O*-methylglucuronic acid from the XOS that have the residue on non-reducing xylosyl residue[119]. Thus, XOS generated from GX by a GH11 need to have the non-reducing xylosyl residue removed by a exo-β-1,4-xylosidase prior to debranching by GH67. In contrast, GH115 has the ability to hydrolyze 4-*O*-methylglucuronic acid residues placed either internal or at the terminal non-reducing end[120].

The function of α-L-arabinofuranosidases in xylan degradation is to remove arabinosyl residues from the xylosyl backbone. These enzymes are grouped into four families; GH43, 51, 54, and 62 and hydrolyze the glycosidic linkage through inversion (GH43) or retention (GH51, and 54). As described in section 1.7 the arabinosyl may be linked to C2, C3 or both OH groups of xylosyl residues. α-L-arabinofuranosidases have been divided into two groups[114]. The major group of these two consists of enzymes active on mono-substituted xylosyl residues on position C2 or C3. This group is disseminated in all of the α-L-arabinofuranosidase families[121]. The minor group includes enzymes active on non-reducing terminal double-substituted xylosyl residues, from which they selectively remove arabinosyl on position C3 leaving the arabinosyl on C2[122]. This group primarily have members from family GH43[122], [123]. Arabinosyl decorations in xylan can be further esterified with ferulic acid or coumaric acid and the CE family catalyzing the de-esterfication belongs to CE1.

The presence of α-D-galactosyl and α-L-galactosyl is more rare and only present in GAX where it contributes to recalcitrance of the xylan[114]. There has not been much research on α-galactosidases that would release the galactosyl residues, but in a study by Rogowski et al they characterize a GH97 and a GH95, which are active on the glycosidic linkage for α-D-galactosyl and α-L-galactosyl, respectively[94].

Acetylxylan esterases are enzymes that hydrolyze the deacetylation of xylan and are present in family CE1-7, 12, and 16. The xylosyl residues in the xylan backbone can be 2-*O* or 3-*O* mono- or double-acetylated. The positional specificity for five of the families has been investigated. CE4 can deacetylate both 2-*O* or 3-*O* acetylations but only if the xylosyl residue is mono-acetylated[124], while CE1, 5, 6, an 16 are capable of deacetylating double-acetylated xylosyl residues[125]. None of the acetylxylan esterases can deacetylate the 3-*O* acetylations on a 4-*O*-methylglucuronic acid substituted xylosyl residue, indicating that 4-*O*-methylglucuronic acid is a steric barrier for esterases[125].

1.14 Enzymatic degradation of mannan

The structure of mannan is heterogeneous (section 1.8) and as with xylan the complete breakdown of mannan requires several enzymes with diverse specificity and mode of action. Mannolytic enzymes are composed of endo-β-1,4-mannanases, exo-β-1,4-mannosidases and β-1,4-glucosidases that hydrolyze the backbone, while α-D-galactosidases and acetylmannan esterases remove the side chain decorations(Figure 12 and Table 7).

For complete degradation of mannan to its constituent sugars these enzymes work together in a concerted manner. Two types of synergies have been identified in relation to mannan degradation; homosynergy and heterosynergy. Homosynergy is defined as cooperativity between two backbone cleaving enzymes (*i.e.* endo-β-1,4-mannanases and exo-β-1,4-mannosidases) or two side chain cleaving enzymes (α-D-galactosidases and acetylmannan esterases), whereas heterosynergy is the synergistic interaction between a side chain cleaving enzyme and a backbone cleaving enzyme (*i.e.* α-D-galactosidases and endo-β-1,4-mannanases)[97], [126].

After xylanases, mannanases are the most important enzymes for the hydrolysis of hemicellulose[95]. Endo-β-1,4-mannanases cleave the β-1,4 glycosidic linkage in the mannan backbone liberating short manno-oligosaccharides (MOS) and have been classified into family GH5, 26, 113 and 134. GH5 and 26 are the largest and best characterized families, while the two other families only have a single or two characterized members respectively. The ability to cleave at the backbone relies for all endo-β-1,4-mannanases on a mannosyl residue at subsite -1, but is also depended on the number and distributions of the side chain decorations[91]. There is significant variation in the specificity of the enzymes, which point out the divergence between GH5 and GH26. GH5 mannanases are able to accommodate mannosyl and glucosyl residues at the -2 and +1 subsites and are thus able to hydrolyze mannosidic linkages flanked by mannosyl and glucosyl residues. In contrast, characterized GH26 mannanases display tight specificity for mannose at both -2 and +1 subsites[127].

Exo-β-1,4-mannosidases are exo-acting enzymes that catalysis the hydrolysis of the β-1,4 glycosidic linkage in mannan and MOS, releasing mannosyl from the non-reducing end. These enzymes have been classified into family GH1, 2, 5, and 113. The presence of a D-galactosyl decorated mannosyl residue adjacent to the terminal mannosyl has been shown to significantly reduce the mannosidase activity compared to a linear substrate[128].

β-1,4-glucosidases are required for the complete degradation of GM, and GaGM. These are exo-acting enzyme that hydrolyze glucosyl residues from the non-reducing end of MOS derived from the degradation of GM, and GaGM by mannanases[97]. The enzymes are classified into family GH1, 3, 9, and 30.

The backbone of GaM and GaGM are decorated with D-galactosyl residues (section 1.8). The α-1,6 galactosidic bonds between D-galactosyl and mannosyl residues are cleaved by α-D-galactosidases. The enzymes fall into family GH4, 27, 31, 36, 97 and 110 and can be further divided into those that can hydrolyze D-galactosyl residues attached to internal mannose residues of oligos or polymers and those that primarily hydrolyze terminal D-galactosyl residues[129].

Acetylmannan esterases hydrolyze deacetylation of *O*-acetyl groups from mannan. Opposed to acetylxylan esterases these enzymes have been long known, however no structures yet exist[129], [130]. In addition, none of the defined families of CE in the CAZy database are ascribed to possess solely acetylmannan activity. This can be due to the fact that some acetylmannan esterases also act as acetylxylan esterases[91].

Table 7 Xylanolytic and mannolytic enzymes

Enzyme	Substrate	Reaction	Family	Clan	Mechanism	Catalytic residues
Endo-β-1,4-xylanase	AX, GX, GAX	Endohydrolysis of 1,4-β-D-xylosidic linkages in xylans	GH5.4	A	R	E,E
			GH5.21	A	R	E,E
			GH8	M	I	D,E
			GH10	A	R	E,E
			GH11	C	R	E,E
			GH30	A	R	E,E
			GH43	F	I	D,E
Exo-β-1,4-D-xylosidase	AX, GX, GAX	Hydrolysis of terminal, non-reducing D-xylose residues in 1,4-β-D-xylooligomers	GH3	-	R	D,E
			GH30	A	R	E,E
			GH39	A	R	E,E
			GH43	F	I	D,E
			GH52	O	R	D,E
			GH120	-	R	D,E
Reducing-end-β-1,4-D-xylosidase	AX, GX, GAX	Hydrolysis of terminal, reducing D-xylose residues in 1,4-β-D-xylooligomers	GH8	M	I	D,E
α-L-arabinofuranosidase	AX, GAX	Hydrolysis of terminal non-reducing α-L-arabinofuranoside residues	GH43	F	I	D,E
			GH51	A	R	E,E
			GH54	-	R	-,-
			GH62	F	-	-,-
α-1,2-D-glucuronidase	GX, GAX	Hydrolysis of terminal α-D-(4-*O*-methyl)glucuronosyl residues	GH67	-	I	-,E
			GH115	-	I	-,-
Acetylxylan esterase	AX, GX, GAX	Deacetylation of xylan	CE1, 3, 5, 7, 12,	-	-	S,H,D
			CE2	-	-	S,H
			CE4	-	-	D,H,H,Co^{2+}/ Zn^{2+}
			CE16	-	-	-
Feruloyl esterase	AX, GAX	Deferuloylation of xylan	CE1	-	-	S,H,D
p-coumaroyl esterase	AX, GAX	De-*p*-coumaroylation of xylan	CE1	-	-	S,H,D
Endo-β-1,4-D-mannanase	M, GaM, GM, GaGM	Endohydrolysis of 1,4-β-D-mannosidic linkages in xylans	GH5	A	R	E,E
			GH26	A	R	E,E
			GH113	A	R	E,E
			GH134	-	I	-,-
Exo-β-1,4-D-mannosidase	M, GaM, GM, GaGM	Hydrolysis of terminal, non-reducing β-D-mannose residues in 1,4-β-D-mannooligos	GH1	A	R	E,E
			GH2	A	R	E,E
			GH5	A	R	E,E
			GH130	-	I	-,-
β-D-glucosidase	GM, GaGM	Hydrolysis of terminal, non-reducing β-D-glucosyl residues with release of β-D-glucose	GH1	A	R	E,E
			GH3	-	R	D,E
			GH9	-	I	D,E
			GH30	A	R	E,E

α-D-galacto-sidase	GaM, GaGM, GAX	Hydrolysis α-D-galactose resi-dues	GH4	-	R	-,-
			GH27	D	R	D,D
			GH31	D	R	D,D
			GH36	D	R	D,D
			GH57	-	R	E,-
			GH97	-	I	E,E
			GH97	-	R	D,E
			GH110	-	I	-,-
α-L-galacto-sidase	GAX	Hydrolysis of terminal, non-reducing α-L-galactose resi-dues	GH95	-	I	N,E
Acetylmannan esterase	GaM, GM, GaGM	Deacetylation of mannan	-	-	-	-,-

R: Retaining, I: Inverting

1.15 Oligosaccharide transporters

Firmicutes utilize a diverse array of transporters to import oligosaccharides for intracellular processing. These transporters fall into three main classes: ATP-binding cassette (ABC) transporters, phosphoenolpyruvate-phosphotransferase system (PTS) transporters, and major facilitator superfamily (MFS) transporters[67]. The first part of this section will focus on ABC importers, which are frequently encoded in glycan utilization loci's together with intracellular GHs in Firmicutes, suggesting that their expression and function is co-regulated[72]. Similar genomic organization of ABC transporters and GHs has also been characterized in other phyla such as Actinobacteria[131]. The second part of this section will introduce how *Bacteriodes* species import oligosaccharides facilitated by analogues to the starch utilization system (Sus).

ABC transporters comprise one of the largest superfamily of membrane proteins[132]. They couple hydrolysis of ATP to translocate substrates across membranes, ranging from ions to macromolecules. Transporters of the ABC type are divided into three functional categories[133]. (I) Importers that mediate the uptake of nutrients including mono- and oligosaccharides (mainly found in prokaryotes). (II) Exporters that are involved in the secretion of various molecules, such as peptides and lipids. (III) The third category of systems is not involved in transport, with some members being involved in translation of mRNA and in DNA repair. All ABC transporters share a characteristic architecture comprising at least two intracellular nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs) that constitute a heterodimeric translocation

pore (Figure 13). The primary structure of TMDs are markedly variable compared to NBDs, which contain highly conserved motifs[132]. The TMDs span the membrane forming a translocation pore that contains a substrate binding site, while the NBDs are molecular motors that transform the chemical potential energy of ATP into protein conformational changes between an outward facing (open to the exterior) and an inward facing (open to cytoplasm) structure of the pore[134]. A "typical" ABC transporter is described to possess 12 transmembrane helices (6 per TMD), however, as stated above, TMDs are heterogeneous and four distinct folds are currently recognized: type I ABC importer, type II ABC importer, type III energy coupling factor importer and ABC exporter[135], [136]. Prokaryotic importers from type I and II, also termed permeases, require an additional high affinity component—the solute binding protein (SBP). These are soluble proteins located in the periplasm of Gram negative bacteria or lipid anchored or fused to the TMD in Gram-positive bacteria and archaea[137]. The role of SBPs is to capture ligands and deliver them to the TMDs, for translocation across the membrane. Based on features from three-dimensional structures Berntsson et al. classified SBPs into six clusters (A-F)[137]. An interesting observation from this classification is that SBPs with similar architecture do not necessarily bind the same ligand. The overall structure of SBPs is composed of two α/β domains connected by a hinge-region and the substrate binding takes place between the two domains[137]. In the absence of ligand the SBP is flexible and exist mainly in an open conformation. Upon substrate binding a closed conformation is stabilized and the ligand trapped. This mechanism is called the Venus Fly-trap model[138]. Based on structural work, a model for the ABC import system has been proposed[139]. In a resting state, the NBD dimer interface is in its inward facing conformation with the two coupling helices farther apart. When SBP interact with the substrate, it undergoes a conformational change from the open to the stabile closed conformation and docks onto the TMDs. This brings the NBD dimer closer such that ATP promotes the outward facing conformation. The substrate will be transferred to the TMD binding site and at the same time position ATP at the catalytic site for hydrolysis. Once ATP is hydrolyzed the TMD will reorient to the inward facing conformation and the substrate diffuse into the cytoplasm.

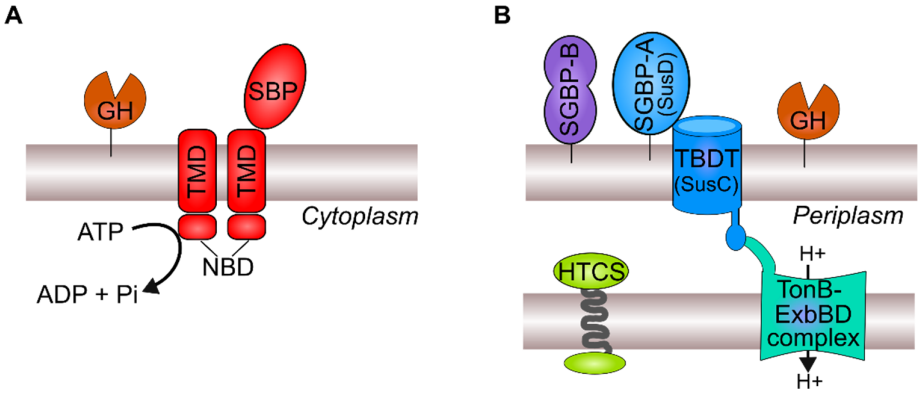


Figure 13 Schematic representation of two bacterial transporter systems. A) Representative structure of an ATP-binding cassette (ABC) transporter composed of a solute binding protein (SBP), two transmembrane domains (TMDs) and two nucleotide-binding domains (NBDs). In Gram positive bacteria and archaea the SBP is lipid anchored or fused to TMD. Translocation of substrates across the membrane is coupled to ATP hydrolysis by NBDs. A glycoside hydrolase (GH) is depicted next the ABC-transporter because long polymers often need to be hydrolyzed to oligosaccharides prior to import. B) Representative structure of a starch utilization system (Sus) in *Bacteroides* species composed of surface glycan-binding proteins (SGBPs), GH(s) and a TonB-dependent transporter (TBDT) in the outer membrane. TBDT interacts with the inner membrane TonB-ExbBD complex consisting of TonB, ExbB and ExbD proteins. Regulation of Sus and the respective GHs is commonly mediated by a hybrid two-component systems (HTCS).

Bacteriodetes have few classical glycan-transporters[140]. Instead their genomes have gene clusters termed polysaccharide utilization loci (PUL) that encode GHs, glycan-binding proteins (SGBPs) and a TonB-dependent transporter(TBDT)[141]. These PUL-encoded proteins work together to selectively facilitate capture and transport of glycans at the cell surface. The Sus from *Bacteriodes thetaiotaomicron* was the first PUL to be characterized and it frequently serves as a model system[140], [142]. Together, eight genes were identified as part of this single gene cluster. All proteins responsible for recognition and initial hydrolysis of starch at the outer membrane, translocation of glycans into the periplasm, further hydrolysis to monosaccharides and transcriptional regulation are encoded in the cluster[143]. Around 18% of all genes in *B. thetaiotaomicron* str. VPI 5482 have been assigned as Sus-like systems because they target other glycans than starch using a similar mechanism as Sus[72]. All mammalian gut Bacteroidetes possess analogous Sus-like systems that typically target numerous diverse glycans[141]. A Sus-like system has at least one sequential pair of *susC* and *susD* homologs[142] that encode TBDT and an N-terminal lipidated SGBP, referred to as SGBP-A[144] (Figure 13). TBDT are intimately associated with SGBP-A that forms a flexible "lid"[145] and deletion of SGBP-A eliminates or reduces PUL function[146], [147]. Glycan binding is often facilitated by one or more structurally distinct SGBPs (e.g., SGBP-B as in figure 13) that may have a specificity identical or complementary to that of SGBP-A[144]. In Sus of *B. thetaiotaomicron* the assisted binding is performed by SusE

and SusF, thus, SGBPs are sometimes referred to as SusE and SusF homologs. Each Sus-like system also contain a repertoire of GHs and other CAZymes. An endo-acting GH hydrolyze glycans at the cell surface and the resulting fragments are actively shuttle into the periplasm by TBDT, where additional linkage specific enzymes act to generate monosaccharides for transport into the cytosol by MFS transporters for catabolism [144].

1.16 Prebiotics

The rich and diverse ecosystem of the HGM is a potential target for alterations that maintain or improve health or treat diseases. The ability to shift the composition and metabolic signature of these microbial populations by rational amendments is now possible through dietary or non-dietary interventions[148]. Dietary fibers represent by far the main source of energy for the HGM and offer opportunities as prebiotics for this beneficial manipulation via diet.

Prebiotics were originally defined by Gibson and Roberfroid in 1995 as "non-digestible food ingredients that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon, and thus improve host health"[149]. The knowledge of the microbial diversity within the gut as well as the relative abundance of different members of the HGM was limited at the time, as it was based almost exclusively on culture- based approaches. Bifidobacteria and lactic acid bacteria, especially lactobacilli, were generally regarded as the main beneficial components of the HGM, but newer molecular methods have since identified a broader range of bacterial taxa having beneficial impact on human health, such as *Roseburia, Eubacterium* or *Faecalibacterium* spp. The prebiotic concept has evolved and the definition been updated several times—latest in 2016. According to the current definition, a prebiotic is a substrate that is selectively utilized by host microorganisms conferring health benefit[148]. Thus, prebiotic effects are no longer limited to the colon but may invoke changes to any host microbial ecosystem.

A number of prebiotic substrates have been reported (Table 8). Among the studied substrates, the best health effect has been reported for fructose-based carbohydrates (inulin and fructooligosaccharides) and galactooligosaccharides[150]. The effect of prebiotics depends not only on the substrate, but the DP of the substrate is also critical for determining the increase in growth of the beneficial bacteria and the site of fermentation.

Members of *Bifidobacterium* preferentially metabolize shorter fructooligosaccharides with a DP of 2-20, while the consumption decreased with DP>20[151]. In a study using a human colon model system the chain length of arabinoxylan-oligosaccharides (AXOS) was found to be critical for the determining the site of fermentation[152]. Shorter AXOS (average DP≤15) were fermented in the ascending and transverse colon, while longer AXOS (average DP=29) reached the distal colon. In the study, the SCFA concentration increased in all regions and the numbers of lactobacilli, *Bacteroides-Prevotella* and *Clostridium coccoides–E. rectale* groups were increased[152]. In general, the carbohydrate gene repertoire, including GHs and transporters, in certain species or strains dictate their differential prebiotic substrate preference based on chain length, monomer constituents, glycosidic linkages, and the overall structural complexity of the oligosaccharides[153]. Ideally, prebiotics would selectively promote beneficial bacterial species, whose populations are decreased in situations associated with an increased disease risk[154]. This clearly requires detailed prior knowledge on the substrate utilization preferences and carbohydrate gene repertoire. Substrate utilization profiles and characterization of carbohydrate gene machinery can be established for cultured strains, providing important information about a specific strain. However, the bacteria in the human gut exist within a complex community, where there is high competition for substrates and cross-feeding of metabolites. This may result in consequences that could not be predicted simply from the substrate preferences, underlining that care must be taken, when transferring results based on *in vitro* experiments to *in vivo* conditions in the evaluation of candidate prebiotics.

Table 8 Selection of reported Prebiotics. Table modified from [155], [156]

Simple prebiotics	Complex prebiotics
Inulin	Pectin
Fructooligosaccharides	Human milk oligosaccharides
Galactooligosaccharides	Resistant starch
Glucooligosaccharides	Arabinoxylan
Isomaltooligosaccharides	Arabinoxylooligosaccharides
Xylooligosaccharides	Polydextrose
Mannooligosaccharides	
Raffinose	
Lactulose	

Chapter 2 Paper 1: Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut

The present paper reports a significant part of the results obtained by during my Ph.D. project. The paper gives insight into xylan catabolism by the butyrate producing human gut symbiont *R. intestinalis*. The enzymatic degradation of xylan and transport of xylan derived oligosaccharides are described in detail in form of a transcriptomic study, enzymatic assays, microscopy, binding studies and NMR. A novel xylan esterase (*Ri*AXE) and a novel xylan binding CBM family were identified. A central finding is that *R. intestinalis*, which was grown on different xylans and xylan components, is able to grow competitively in co-culture with the model xylan degrading *Bacteroides ovatus*. Strikingly, *R. intestinalis* was shown to possess a preference for oligomers of 4–5 xylosyl units, in contrast to the *B. ovatus* competitor, which targets larger ligands. This differential preference was illustrated by the out-competition of *B. ovatus* in a mixed co-culture using a preferred ligand for the *R. intestinalis* ABC importer. I have aside from the NMR and expression + characterization of the transporter, performed and analyzed the all experiments in collaboration with the co-authors.

Supplementary information is included in the end of the chapter

ARTICLES
https://doi.org/10.1038/s41564-018-0132-8
nature microbiology

Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut

Maria Louise Leth[1], Morten Ejby[1], Christopher Workman[1], David Adrian Ewald[1], Signe Schultz Pedersen[1], Claus Sternberg[1], Martin Iain Bahl[2], Tine Rask Licht[2], Finn Lillelund Aachmann[3], Bjørge Westereng[4] and Maher Abou Hachem[1*]

Metabolism of dietary glycans is pivotal in shaping the human gut microbiota. However, the mechanisms that promote competition for glycans among gut commensals remain unclear. *Roseburia intestinalis*, an abundant butyrate-producing Firmicute, is a key degrader of the major dietary fibre xylan. Despite the association of this taxon to a healthy microbiota, insight is lacking into its glycan utilization machinery. Here, we investigate the apparatus that confers *R. intestinalis* growth on different xylans. *R. intestinalis* displays a large cell-attached modular xylanase that promotes multivalent and dynamic association to xylan via four xylan-binding modules. This xylanase operates in concert with an ATP-binding cassette transporter to mediate breakdown and selective internalization of xylan fragments. The transport protein of *R. intestinalis* prefers oligomers of 4–5 xylosyl units, whereas the counterpart from a model xylan-degrading *Bacteroides* commensal targets larger ligands. Although *R. intestinalis* and the *Bacteroides* competitor co-grew in a mixed culture on xylan, *R. intestinalis* dominated on the preferred transport substrate xylotetraose. These findings highlight the differentiation of capture and transport preferences as a possible strategy to facilitate co-growth on abundant dietary fibres and may offer a unique route to manipulate the microbiota based on glycan transport preferences in therapeutic interventions to boost distinct taxa.

The human gut microbiota (HGM) is recognized as a determinant of human health and metabolic homeostasis[1,2]. Specific signatures of the HGM are associated with local and systemic disorders, including irritable bowel disease, obesity, type 2 diabetes and colon cancer[3]. The composition of the HGM is greatly affected by non-digestible dietary glycans[4,5]. Only a few species of the HGM are equipped to deconstruct and ferment distinct complex glycans into short-chain fatty acids (SCFAs)[6]. The effect of SCFAs on host health and physiology remains an important aspect of the microbiota–host interaction. The SCFA butyrate, the preferred energy source for colonocytes, possesses anti-inflammatory roles and reduces the risks of colon cancer and enteric colitis[7–10]. Butyrate-producing Firmicutes are abundant in healthy individuals, but are markedly reduced in patients with inflammatory disorders[11,12]. Butyrate producers including *Roseburia* spp. are increased in patients with metabolic syndrome after faecal transfer therapy and correlate positively to improvement of insulin resistance[13]. Investigations of the metabolic preferences of butyrate producers and their interplay with major HGM commensals are instrumental to develop interventions targeting butyrate deficiency-related disorders.

Roseburia is a common genus of *Clostridium* cluster XIVa within Firmicutes that harbour prevalent butyrate producers[14,15]. This taxon adheres to mucin, reflecting intimate association with the host[16]. *Roseburia intestinalis* strains encode an impressive repertoire of carbohydrate-active enzymes (CAZymes) compared to most Firmicutes[17]. *R. intestinalis*, the taxonomically related *Eubacterium rectale*, and species from *Bacteroides* are the only known HGM taxa that utilize the hemicellulose xylan[18–20]. Xylan is particularly abundant in cereal grains (arabinoxylan) but is also found in fruits and vegetables (glucuronoxylan)[21] (Fig. 1a). Xylan utilization by *Bacteroides*-dominant gut commensals has been investigated in detail[22,23], but similar knowledge is lacking for Firmicutes counterparts.

Here, we show that *R. intestinalis* L1-82 grows on dietary-relevant xylans, with a preference for cereal arabinoxylans. The growth is mediated by a multi-modular cell-attached xylanase and by an ATP-binding cassette (ABC) transporter. We have characterized the xylanolytic enzymes and the transport protein, which enabled modelling xylan utilization by *R. intestinalis* and the identification of two previously undescribed xylan-specific CAZyme families. *R. intestinalis* efficiently competes with a model xylan degrader from *Bacteroides* when grown on soluble and insoluble xylans. Strikingly, transport proteins that confer xylo-oligosaccharides capture in *R. intestinalis* and *Bacteroides* targeted ligands of different sizes, thus markedly reducing the competition for preferred ligands by either taxon. These results emphasize the competitiveness of butyrate-producing Firmicutes in targeting key dietary fibres such as xylan and highlight differential glycan capture and transport as an important feature in co-growth on abundant dietary fibres such as xylan.

Results

Inducible cell-attached xylanase activity mediates growth of *R. intestinalis* on substituted xylans. Anaerobic growth of *R. intestinalis* L1-82 was measured on soluble and insoluble xylans (Fig. 1b–d). *R. intestinalis* L1-82 grows rapidly on soluble xylans, especially wheat

[1]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark. [2]National Food Institute, Technical University of Denmark, Lyngby, Denmark. [3]NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway. [4]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway. *e-mail: maha@bio.dtu.dk

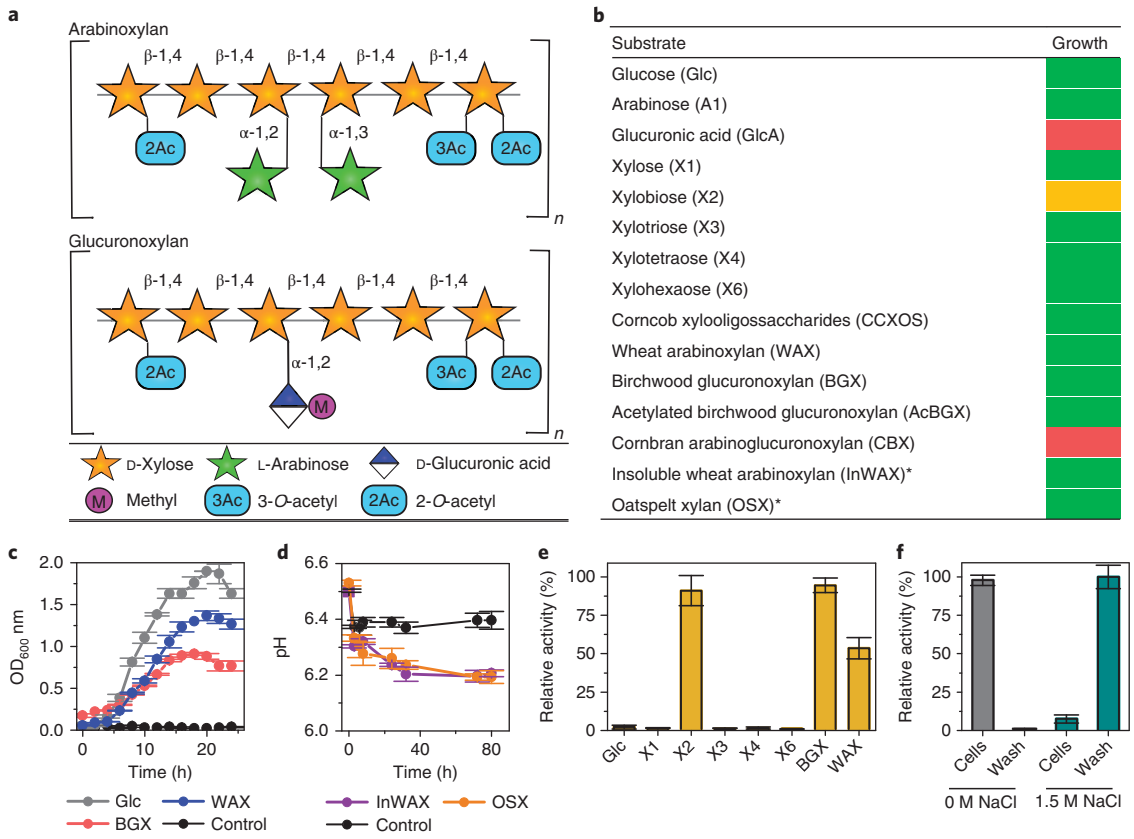


Fig. 1 | Growth of *R. intestinalis* and induction of extracellular activity. a, Schematic representation of cereal arabinoxylan and glucuronoxylan present in the dicot cell wall, for example, in fruits and vegetables. **b**, Growth level for 18 hours on xylans, oligosaccharide and monosaccharides thereof, with glucose as a control. Green: OD_{600} increase of >1.0 for soluble substrates and a pH drop of >0.3 for insoluble xylans; yellow: $0.3 < \Delta OD_{600} < 0.5$; red: $\Delta OD_{600} < 0.1$. Asterisks indicate insoluble xylans. Data are means of triplicates. **c**, Growth curves on glucose (Glc), WAX, BGX and a no-carbon-source control. **d**, Growth on InWAX and oatspelt xylan (OSX) and no-carbon-source control. All growth measurements (**c**,**d**) are means of triplicates with standard deviations. **e**, Xylanase activity of *R. intestinalis* cells grown on glucose, xylo-oligosaccharides, BGX and WAX for 18 hours. **f**, Cells grown on BGX were washed (PBS buffer ± 1.5 M NaCl) and xylanase activity was measured in wash and cell fractions to verify the localization of the enzymes. Xylanase activity (**e**,**f**) was measured using the DNS-reducing sugar assay, and data are triplicates with standard deviations.

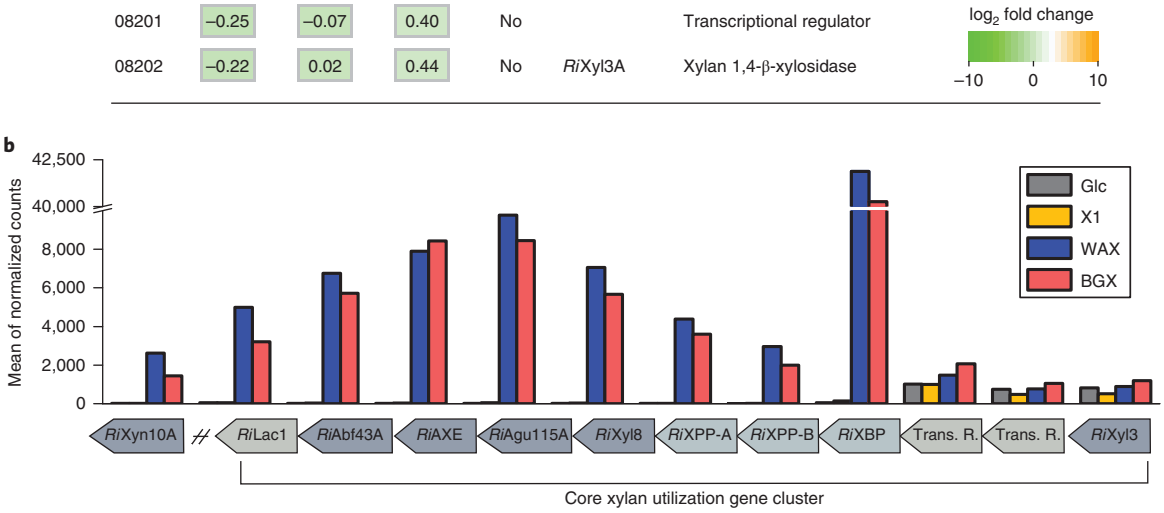
arabinoxylan (WAX; $\mu_{max} = 0.26\,h^{-1}$), compared to birch glucuronoxylan (BGX; $\mu_{max} = 0.13\,h^{-1}$) (Fig. 1c). This bacterium also utilizes highly acetylated xylans and insoluble cereal arabinoxylans from wheat (InWAX) and oat spelt, but not cornbran glucuronoarabinoxylan (CBX). Xylo-oligosaccharides and xylan-derived monosaccharides (except glucuronic acid) were also utilized. Extracellular endo-1,4-β-xylanase (hereafter xylanase) activity was induced upon growth on BGX, WAX and the poor growth substrate xylobiose (X2) (Fig. 1e). Xylanase activity was cell attached, but was released in the presence of high-salt concentration (Fig. 1f), suggesting non-covalent attachment.

Genes encoding an ABC transporter and a multi-modular xylanase are top upregulated in response to growth on xylan. We performed an RNA sequencing (RNA-seq) transcriptional analysis of *R. intestinalis* grown on WAX, BGX, xylose and glucose. Of the 4,777 predicted genes, 1–3.5% were highly upregulated ($\log_2$ fold change > 5) on xylans compared to glucose (Supplementary Table 1) and the majority were involved in carbohydrate metabolism. Besides a separate locus encoding a multi-modular xylanase of glycoside hydrolase family 10 (GH10 according to CAZy classification[24]), the top genes in the xylan transcriptomes cluster on a single locus (Fig. 2a,b) that contains 11 genes, including 4 xylanolytic CAZymes: GH43, GH115, GH8 and GH3. Only one (ROSINTL182_08192, LacI type) of three transcriptional regulators was highly upregulated. Strikingly, the most upregulated gene in the xylan transcriptomes encodes a solute-binding protein (SBP) of an ABC transporter. The permease genes of this transporter were among the top six upregulated by xylans. Signal peptides were only predicted for the xylanase and the transporter SBP, which is consistent with the extracellular breakdown of xylan followed by capture and uptake of xylo-oligosaccharides by the ABC transporter. The expression and the localization of the SBP and the xylanase at the cell surface were corroborated using immunofluorescence microscopy (Fig. 2c). Two additional loci, unique to the *R. intestinalis* L1-82 strain, were also upregulated albeit markedly less (Supplementary Fig. 1a–d). One of these loci encodes a second active GH10 xylanase, which is expressed at the cell surface (Supplementary Fig. 1e). The transcriptomic analyses enabled the assignment of the xylose ABC transporter and the genes involved in intracellular metabolism of xylose, arabinose and glucuronic acid (Supplementary Fig. 2).

A previously undescribed family of binding modules confers extended and dynamic xylan binding to the multi-modular xylanase. The highly upregulated *Ri*Xyn10A, which is conserved within *R. intestinalis* (Supplementary Fig. 3), is one of the largest known xylanases from the HGM (Supplementary Table 2). *Ri*Xyn10A comprises an amino-terminal unassigned domain (residues 28–165), a CBM22 xylan-binding module, a GH10 catalytic module, two tandem CBM9 xylan-binding modules, a bacterial immunoglobulin-like

a

Locus ID	log_2 fold change			SP	Protein	Annotation
	X1/Glc	WAX/Glc	BGX/Glc			
06494	−1.04	7.00	6.17	Yes	*Ri*Xyn10A	Endo-1,4-β-xylanase
08192	−0.46	6.46	5.82	No	*Ri*Lac1	Transcriptional regulator, LacI family
08193	0.04	8.69	8.45	No	*Ri*Abf43A	α-L-Arabinofuranosidase
08194	0.24	8.78	8.88	No	*Ri*AXE	Acetyl xylan esterase
08195	0.63	8.55	8.35	No	*Ri*Agu115A	Xylan α-1,2-glucuronidase
08196	0.78	8.60	8.29	No	*Ri*Xyl8	Reducing-end xylose-releasing exo-oligoxylanase
08197	0.91	8.71	8.43	No	*Ri*XPP-A	ABC transporter, permease protein
08198	0.03	8.89	8.33	No	*Ri*XPP-B	ABC transporter, permease protein
08199	0.49	9.12	9.07	Yes	*Ri*XBP	ABC transporter, xylan-binding protein
08200	0.28	0.46	0.96	No		Transcriptional regulator
08201	−0.25	−0.07	0.40	No		Transcriptional regulator
08202	−0.22	0.02	0.44	No	*Ri*Xyl3A	Xylan 1,4-β-xylosidase



c

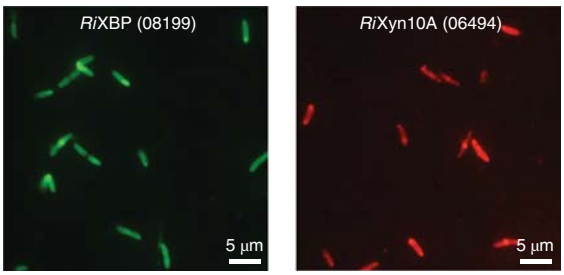


Fig. 2 | The core xylan utilization apparatus of *R. intestinalis*. **a**, The RNA-seq heatmap depicts log_2 fold changes of the top upregulated xylan utilization genes expressed by cells grown on xylose (X1), WAX and BGX relative to glucose. Experiments are performed in biological triplicates, and locus tag numbers ROSINTL182_xxxxx are abbreviated with the last numbers after the underscore. **b**, Gene expression depicted as the mean of the normalized Deseq2 gene counts for the core xylan utilization genes shown in **a**. Trans. R., transcriptional regulator; SP, signal peptide. **c**, Representative micrographs showing the extracellular localization of *Ri*XBP and *Ri*Xyn10A, the SBP of the xylo-oligosaccharide-specific ABC transporter and the xylanase, respectively, were visualized by fluorescence microscopy of *R. intestinalis* cells in biological triplicates using primary antibodies against these two proteins.

domain group 2 (pfam02368)[25] and a *Listeria-Bacteroides* repeat domain (pfam09479)[26]. The two latter positively charged domains (residues 1,100–1,356; pI > 10) (Fig. 3a) probably mediate attachment of the enzyme to the cell[25–27].

*Ri*Xyn10A incubated with BGX, WAX and InWAX generated linear and decorated oligosaccharides (Fig. 3b,c), but was inactive on highly and heterogeneously substituted arabinoglucuronoxylan from corn bran, consistent with the growth profile (Fig. 1b). The enzyme was inactive on X2 and showed poor activity on xylotriose (X3) (Supplementary Fig. 4a). Xylotetraose (X4) and xylopentaose (X5) were efficiently hydrolysed, revealing the requirement for the occupancy of at least four substrate-binding subsites for high turnover.

A BLASTp search of the N-terminal domain (previously undescribed carbohydrate-binding module (CBMx)) against UniProt gave no hits indicating the lack of homologues with assigned function. CBMx confers affinity to xylan based on a two-times higher Michaelis constant (K_m) when CBMx was deleted from *Ri*Xyn10A

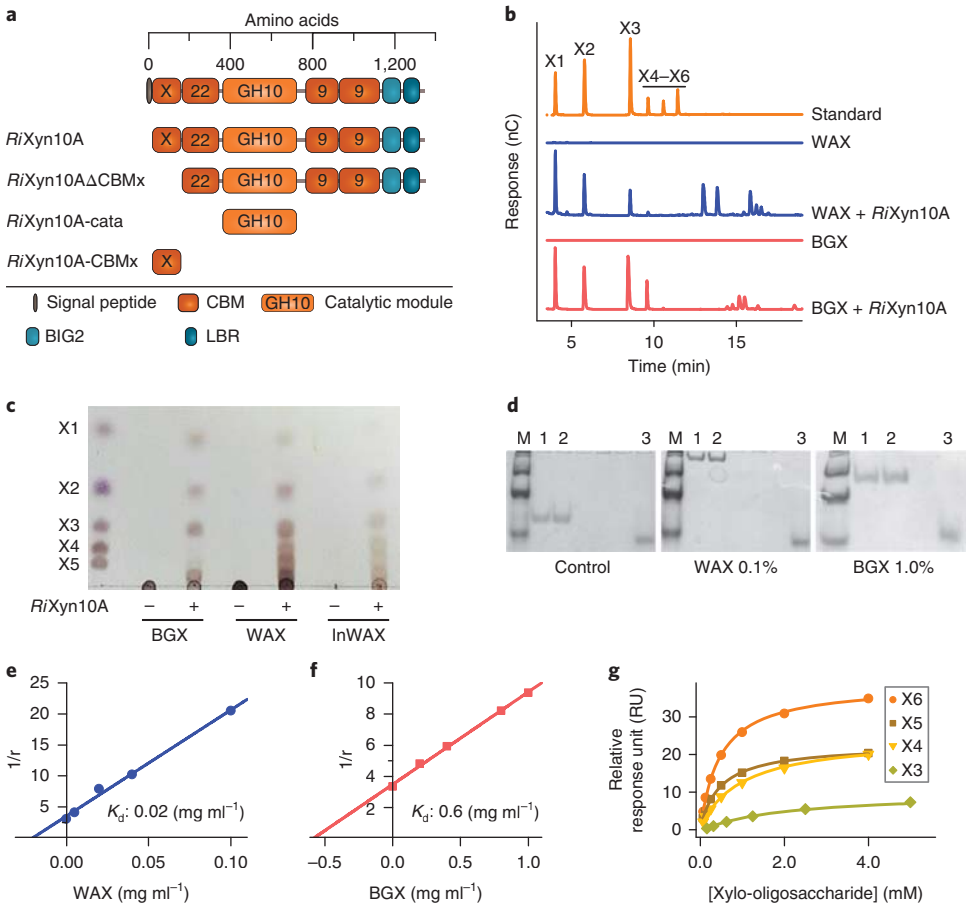


Fig. 3 | A low-affinity xylan-binding module mediates extended xylan binding to the xylanase *Ri*Xyn10A. a, Domain organization of *Ri*Xyn10A and truncated variants, all produced recombinantly without the native signal peptide. BIG2, bacterial immunoglobulin-like domain group 2; CBM, carbohydrate-binding module; CBMx, previously undescribed CBM; LBR, *Listeria-Bacteroides* repeat domain. **b**, Xylanase activity of *Ri*Xyn10A on WAX and BGX assayed by HPAEC-PAD. The peaks eluting after X6 are probably decorated xylo-oligosaccharides. **c**, *Ri*Xyn10A activity analysed using thin layer chromatography. The – and + indicate absence or presence of *Ri*Xyn10A, respectively. **d**, Binding of *Ri*Xyn10A-CBMx to the negative control (no polysaccharide), WAX or BGX xylans analysed using affinity electrophoresis. The uncropped gels are in Supplementary Fig. 4h. Lanes 1+2, *Ri*Xyn10A-CBMx (3.0 μg); lane 3, β-lactoglobulin negative control (1.5 μg); M, marker. **e**,**f**, Binding data from **d** is depicted as a plot of 1/*r* against xylan concentration, where r is the relative migration distance of *Ri*Xyn10A-CBMx in the presence of xylans in the gel. The K_d from this analysis is shown in each plot. **g**, Binding isotherms of *Ri*Xyn10-CBMx to xylo-oligosaccharides. Solid lines are fits of a one-binding site model to the SPR sensograms. Experiments were performed in duplicates, except in **g**, which were performed in triplicates.

(Table 1 and Supplementary Fig. 4b–g). Affinity electrophoresis established CBMx as a xylan-binding module and revealed a 30-fold higher dissociation constant (K_d) on WAX compared with BGX (Fig. 3d–f and Supplementary Fig. 4h). Surface plasmon resonance (SPR) analysis revealed the highest affinity towards xylohexaose (X6) (Table 2, Fig. 3g and Supplementary Fig. 5a–e). The relatively low binding affinity to X6 ($K_d \approx 0.5$ mM) was corroborated using isothermal titration calorimetry (ITC) (Table 2, Supplementary Fig. 5f and Supplementary Table 3). Deleting CBMx decreased the average K_d of *Ri*Xyn10A from 128 μM to 65.4 μM (*Ri*Xyn10AΔCBMx) (Table 2 and Supplementary Fig. 5g–j), asserting that one or more of the other CBMs possess a higher affinity compared to the N-terminal module. Homologues (sequence identity: 55–27%) of CBMx are present mainly in other *Clostridium* XIVa cluster Firmicutes (Supplementary Table 4).

Table 1 | Xylan hydrolysis kinetics of *Ri*Xyn10A and truncated variants

Substrate	Enzyme	K_m (mg ml^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (ml mg^{-1} s^{-1})
WAX	*Ri*Xyn10A	2.4 ± 0.4	202 ± 14	84
	*Ri*Xyn10AΔCBMx	5.5 ± 1.2	386 ± 47	71
	*Ri*Xyn10A-cata	ND	ND	34
BGX	*Ri*Xyn10A	3.3 ± 0.7	196 ± 18	59
	*Ri*Xyn10AΔCBMx	6.6 ± 1.6	369 ± 45	56
	*Ri*Xyn10A-cata	ND	ND	16

Kinetics of the *Ri*Xyn10A-cata are not modelled by the Michaelis–Menten expression, and catalytic efficiencies are estimated from the linear regression of the initial rate data. Data are means of triplicates with standard deviations. ND indicates low substrate affinity precluding determination of the kinetic parameters.

Preference of the binding protein of the ABC transporter that mediates uptake of xylo-oligosaccharides. Complex xylo-oligosaccharides decorated with arabinosyl and 4-*O*-methyl-glucuronosyl are produced by *Ri*Xyn10, based on the decrease levels in some of these products and the increase levels in arabinose, 4-*O*-methyl-glucuronic acid (MeGlcA) and un-substituted xylo-oligosaccharides after treatment with debranching enzymes (see the next section). Transcriptional analysis (Fig. 2a) identified a

Table 2 | Binding parameters of the xylan-binding module CBMx and RiXyn10A and variants

Variant	Ligand	K_d (μM)
*Ri*Xyn10A-CBMx	WAX	0.02[a]
	BGX	0.6[a]
	X6	479 ± 26
	X6	413 ± 125[b]
	X5	490 ± 15
	X4	998 ± 42
	X3	1,900 ± 220
	Man6	ND
*Ri*Xyn10A	X6	128 ± 7.1
*Ri*Xyn10AΔCBMx	X6	65.4 ± 8.4

K_d determined by SPR are the means of a duplicate with the standard deviations. [a]K_d (mg ml^{-1}) was determined by affinity electrophoresis. [b]K_d was determined by ITC. The SPR experiments are performed in triplicates and the data are the means with standard deviations. The affinity electrophoresis and the ITC data are from single experiments. The error estimate in the ITC experiment is from the fit of a one-binding site model to the binding isotherm.

Table 3 | The preference of the xylo-oligosaccharide transport protein from *R. intestinalis*

Ligand	K_d (μM)	N_0	ΔH (kcal per mol)	$T\Delta S$ (kcal per mol)	ΔG (kcal per mol)
X6	112.7	1.19	−9.01	−3.6	−5.4
X5	10.3	0.86	−13.54	−6.7	−6.8
X4	16.5	0.68	−12.8	−6.3	−6.5
X3	225.7	0.58	−21.1	−16.1	−5.0
X2	ND				
AX3	215.5	0.26	−44.3	−39.4	−4.9
AX4	6.8	0.58	−12.3	−7.0	−5.3

Binding energetics of the transport protein *Ri*XBP to xylo-oligosaccharides determined by ITC. Data are the means of duplicate experiments. N_0, ΔH, $T\Delta S$ and ΔG denote the binding stoichiometry, enthalpy, entropy and Gibbs free energy, respectively. ND indicates affinity too low to be determined. The cartoons to the right side depict the analysed susbtrates using the same code as in Fig. 1a, and the chemical structures of arabinosyl-decorated oligomers AX3 and AX4 are shown in Supplementary Fig. 6h,i.

plausible ABC transporter of xylo-oligosaccharides. The preference of SBPs associated with oligosaccharide-specific ABC transporters correlates with the bacterial uptake preference[28,29]. We measured the affinity of *Ri*XBP, the SBP of the top upregulated ABC transporter on xylo-oligomers (Table 3 and Supplementary Fig. 6a–i). The preferred un-substituted ligand was X5 followed by X4, and the affinity decreased steeply for smaller or larger oligosaccharides. Internal arabinosyl decorations (AX4) were preferred based on the 2.4-times higher affinity compared to X4. The tolerance and recognition of arabinosylated ligands are supported by the good growth on WAX. In summary, *Ri*XBP is selective in capturing branched xylo-oligosaccharides with preference for a backbone of 4–5 xylose residues.

The concentrations of oligosaccharides in spent supernatants from *R. intestinalis* cultures during growth on xylan were too low for reliable detection using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD), which suggests an efficient uptake of oligomeric products (Supplementary Fig. 6j,k).

***R. intestinalis* degrades internalized decorated xylo-oligosaccharides by the concerted action of three hydrolases and a previously undescribed family of acetyl esterases**. Xylo-oligosaccharides are degraded in the cytoplasm after uptake. We characterized the α-glucuronidase *Ri*Agu115A (GH115), the α-L-arabinofuranosidase *Ri*Abf43A (GH43), two xylosidases *Ri*Xyl8 (GH8) and *Ri*Xyl3A (GH3) as well as *Ri*AXE (ROSITNL182_08194; GenBank accession EEU99941.1) from the core xylan utilization locus.

*Ri*Agu115A released MeGlcA from glucuronoxylans (BGX and beechwood glucuronoxylan) and from BGX pretreated with *Ri*Xyn10A (Fig. 4a,c–f and Supplementary Fig. 7a). The k_{cat}/K_m of *Ri*Agu115A was 16-fold higher on glucuronoxylan hydrolysate compared to intact glucuronoxylan (Supplementary Table 5), consistent with the intracellular enzyme localization. This enzyme also cleaves MeGlcA decorations at the xylosyl penultimate to the reducing end (Supplementary Fig. 7b,c), but its activity was blocked by acetylations (Fig. 4d). The released MeGlcA is metabolized intracellularly in agreement with the transcriptomic data (Supplementary Fig. 2a). The lack of growth on glucuronic acid (Fig. 1b) is probably due to the lack of an uptake system for this monosaccharide.

*Ri*Abf43A is an α-L-arabinofuranosidase that exclusively releases arabinose from WAX (Fig. 4a and Supplementary Fig. 7d). Kinetic analysis against WAX and AX4 (Supplementary Table 6) revealed recognition of internal arabinosyl substitutions. Degradation of WAX by xylanases results in the production of significant amounts of double-arabinosyl-substituted oligomers. The lack of accumulation of these to detectable levels (Supplementary Fig. 6j) is suggestive of their uptake and debranching by *Ri*Abf43A or other upregulated arabinofuranosidases.

Both *Ri*Xyl3A and *Ri*Xyl8 generated xylose from xylo-oligosaccharides, but lacked activity towards xylan (Supplementary Tables 7,8 and Supplementary Fig. 7d–g). *Ri*Xyl3A degraded xylo-oligosaccharides to xylose, whereas *Ri*Xyl8 was inactive towards X2. Reduction of xylo-oligosaccharides abolished the activity of *Ri*Xyl8, assigning it as a reducing-end β-xylosidase[30] (Supplementary Fig. 7h), in contrast to *Ri*Xyl3A that recognizes non-reducing xylosyl units. The concerted and overlapping activities of these enzymes (Supplementary Fig. 7) result in rapid depolymerization of arabinosyl- and MeGlcA-decorated xylo-oligosaccharides. Induction of extracellular xylanase activity by X2, but not by X3–X6, is intriguing as X2 is transiently formed during cytoplasmic degradation of X3–X6. A possible explanation is that the induction is mediated by X2 binding to an extracellular regulatory protein, similar to the xylose sensory system in clostridial Firmicutes[31].

*Ri*AXE, which lacks functionally described homologues, was highly upregulated on xylans (Fig. 2a). This enzyme shares conserved residues with the SGNH lipase esterase superfamily (Pfam cd00229), but possesses low sequence identities to esterases from CAZy. Homologues of this enzyme are encoded by *Clostridium* cluster XIVa strains from the human gut and a range of Firmicutes (Supplementary Fig. 8a). Assaying *Ri*AXE activity towards acetylated BGX (AcBGX) oligosaccharides (generated with *Ri*Xyn10A) using NMR revealed efficient deacetylation of both 2-*O*-acetylated xylose (C2) and 3-*O*-acetylated xylose (C3), but with a preference for C2 decorations (Fig. 4b, Supplementary Tables 9,10 and Supplementary Fig. 8). *Ri*AXE left a single acetyl group on the AcBGX oligosaccharides (Fig. 4e). Inclusion of *Ri*Agu115A in this reaction resulted in complete deacetylation (Fig. 4f), suggesting that close MeGlcA decorations protect closeby acetylations. Analysis of the deacetylation rates unveiled the concerted action with *Ri*Agu115A and preference to hydrolysates of *Ri*Xyn10A rather than intact xylan (Supplementary Fig. 8b,c). *Ri*AXE specifically recognizes acetylations on xylosyl units based on the lack of activity on acetylated chitin and very low activity on acetylated mannan and cellulose monoacetate (Supplementary Table 11). Accordingly, *Ri*AXE is an efficient xylan-specific representative of a previously undescribed acetyl esterase family.

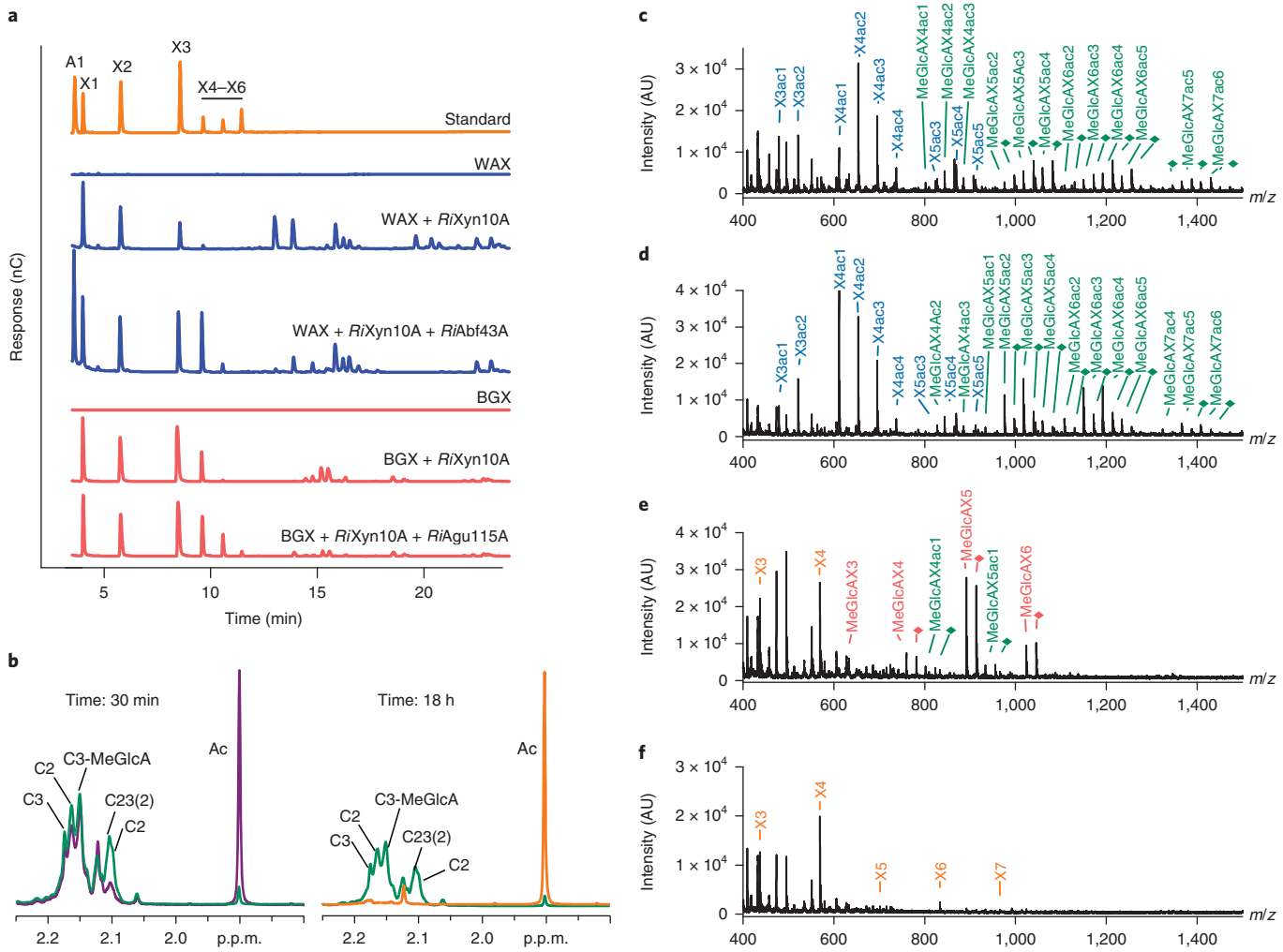


Fig. 4 | Intracellular xylo-oligosaccharide depolymerization. a, α-L-Arabinofuranosidase and α-glucuronidase activity on WAX and BGX for *Ri*Abf43A and *Ri*Agu115A, respectively, based on HPAEC-PAD analysis. **b**, Time-resolved NMR for *Ri*AXE deacetylation of AcBGX treated with *Ri*Xyn10A and *Ri*Agu115A. Deacetylation time course for the first 30 min and after 18 h (green: 0 min, purple: 30 min and orange: 18 h). All verified signals with 2-*O*-acetylation decreased faster in the initial phase of the reaction. The proton spectra of the acetylated region show nearly complete deacetylation after 18 h. The signal at 2.13 p.p.m. is probably attributed to another acetylated sugar residue. Acetyl groups are designated as: C2; C3; C23, 2,3-di-*O*-acetylated xylose; C3-MeGlcA, 4-*O*-methyl glucuronic acid 2-*O*-substituted and 3-*O*-acetylated xylose; C23(2), signal for the 2-*O*-acetylated of C23. The assignment of the acetylated sugar signals was based on homonuclear and heteronuclear NMR correlation experiments (Supplementary Fig. 8). **c**–**f**, Hydrolysis products from AcBGX by *Ri*Xyn10A (**c**), *Ri*Xyn10A and *Ri*Agu115A (**d**), *Ri*Xyn10A and *Ri*AXE (**e**), and *Ri*Xyn10A, *Ri*Agu115A and *Ri*AXE (**f**). Enzyme action was analysed by MALDI-TOF mass spectrometry. Xylo-oligosaccharides decorated with acetyl and MeGlcA are in green, acetyl in blue, MeGlcA in red and no sidechains in orange. Di-sodium adducts of a MeGlcA-decorated oligosaccharides (diamonds) are coloured as their corresponding single sodium adducts. Experiments were performed in duplicates, except for the NMR, where they were performed once. AU, arbitrary units.

In summary of the biochemical characterization presented above, we propose a model for the uptake and degradation of diet-derived acetylated arabinoxylan and glucuronoxylan by *R. intestinalis* L1-82 (Fig. 5a).

***R. intestinalis* competes with *Bacteriodes* for xylans**. The growth potential of *R. intestinalis* was compared to the efficient xylan-degrader *Bacteroides ovatus*[22], by observing the growth of individual cultures and in co-culture. Both strains displayed similar growth on xylan as a carbon source (Fig. 5b–d and Supplementary Fig. 9a,b). In competition, both strains appeared to grow well on xylans, whereas *R. intestinalis* dominated the co-culture on X4 after 7 hours of growth (Fig. 5e–h). The results indicate that *R. intestinalis* is an efficient primary degrader of xylan that is able to compete with *B. ovatus* and outcompete this bacterium on preferred smaller xylo-oligosaccharides.

Discussion

The human gut is dominated by bacteria from the Firmicutes and Bacteriodetes phyla. Firmicutes are regarded as metabolic specialists, whereas Bacteroidetes (mainly *Bacteroides*) are considered generalists based on narrow versus broad glycan utilization capabilities, respectively[6]. The size and diversity of encoded CAZymes reflect these metabolic labels, which applies to *R. intestinalis*, based on the limited glycan growth profiles[5]. However, this species possesses distinctively larger CAZymes than most known gut clostridial Firmicutes[17]. *R. intestinalis* has been proposed to be a key xylan degrader in the human gut along with specific species of *Bacteroides*[18,19]. Enumeration of *R. intestinalis* on xylans including wheat bran is reported in vitro and in vivo[20,32]. However, insight is lacking on the preferences and the molecular machinery evolved by *R. intestinalis* to target xylan compared to *Bacteriodes* members. In this study, we present a model underpinning the molecular basis

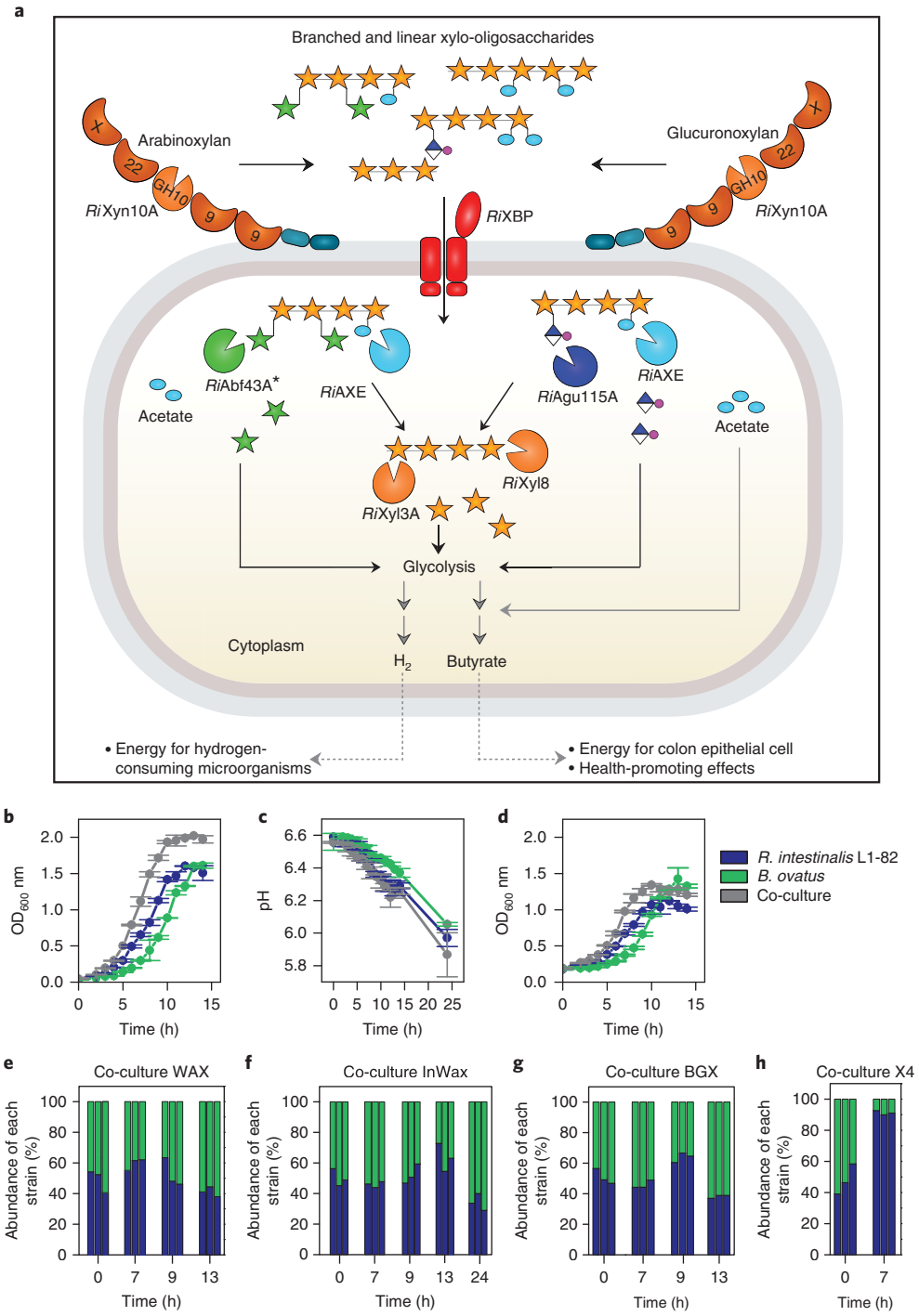


Fig. 5 | Model for xylan utilization by *R. intestinalis* and competition assay with *B. ovatus*. a, *Ri*Xyn10A on the cell surface efficiently captures diet-derived acetylated arabinoxylan and acetylated glucuronoxylan by its CBMs and hydrolyses it into linear and decorated xylo-oligosaccharides, which are subsequently captured by *Ri*XBP for uptake into the cytoplasm. Internalized xylo-oligosaccharides are debranched and hydrolysed into monosaccharides and acetate. Xylose and arabinose are converted to xylulose 5-phosphate before entering the pentose phosphate pathway, whereas MeGlcA is converted to 2-oxo-3-deoxygalactonate 6-phosphate. These precursors enter glycolysis, which generates pyruvate, some of which is converted to butyrate[54]. The asterisk next to *Ri*Abf43A indicates activity on both α-1,2-linked and α-1,3-linked L-arabinose. The black solid arrows show steps established or confirmed in this study. The grey solid arrows indicate steps for the production of H_2 and butyrate by *R. intestinalis*[15], which has been demonstrated for this species during xylose fermentation[19]. The grey dashed arrows indicate that H_2 and butyrate are externalized by unknown mechanisms. To make the model more general for the *R. intestinalis* species, the second less upregulated extracellular xylanase *Ri*XynB, which is unique for the L1-82 strain, is not included in the model, although it is expressed at the cell surface. The figure keys in Figs. 1a and 3a also apply here. **b**–**d**, Growth of the monoculture and co-cultures of *R. intestinalis* and *B. ovatus* on WAX (**b**), InWAX (**c**) and BGX (**d**). Data are the means of a biological triplicate with standard deviations. **e**–**h**, Time course of the relative strain abundance during growth of co-cultures on xylans and X4 determined by quantitative PCR. All data are the means of a biological triplicate.

for xylan utilization by *R. intestinalis* L1-82 as a representative for prevalent butyrate-producing clostridia (Fig. 5a). *R. intestinalis* is a primary degrader equipped with a highly efficient machinery for the utilization of complex dietary xylans, including insoluble arabinoxylan from cereals. Key components of the *R. intestinalis* xylan utilization strategy include a multi-modular extracellular xylanase and an ABC transporter, which confer the capture, breakdown and internalization of decorated xylo-oligosaccharides. In the cytoplasm, internalized oligomers are degraded without loss to competing species. *R. intestinalis* grows on acetylated xylan, which reflects an adaptation to this abundant decoration in dietary xylans (Fig. 1b). Internalized acetylated xylo-oligosaccharides are metabolized by an intracellular previously unknown esterase family that is capable of removing C2, C3 and double acetylations (Fig. 4b and Supplementary Fig. 8).

The extracellular xylanase *Ri*Xyn10A, the ABC transporter and enzymes conferring cytoplasmic breakdown of xylan oligosaccharides were assigned as the core xylan utilization apparatus of *R. intestinalis* (Fig. 2a,b). This was based on (1) conservation of this apparatus within this species (Supplementary Fig. 3), (2) the highest transcriptional upregulation (Fig. 2a), and (3) biochemical data from the present study. The two additional xylan-upregulated loci in *R. intestinalis* L1-82 (Supplementary Fig. 1) are lacking in the xylan-utilizing *R. intestinalis* XB6B4 and *R. intestinalis* M50/1 (ref. [20]). The expression and activity of the xylanase *Ri*Xyn10B, encoded by an auxiliary locus (Supplementary Table 12 and Supplementary Fig. 1c,e), supports the participation of more than one locus in xylan breakdown in *R. intestinalis* L1-82. Multiplicity of xylan utilization loci is implicated in targeting a larger structural diversity of naturally occurring xylans by *Bacteroides*[22], which may also apply for *R. intestinalis*.

The *R. intestinalis* core xylanase *Ri*Xyn10A mediates the capture and breakdown of arabinoxylan and glucuronoxylan (Figs. 1,3). This enzyme possesses four CBMs from two known and one previously undescribed xylan-binding families, representing the most complex organization of HGM xylanases (Fig. 3a and Supplementary Table 2). This organization is conserved within currently sequenced *R. intestinalis* species, whereas other *Clostridium* XIVa taxa possess simpler enzymes that lack one or more of the *Ri*Xyn10A CBMs. The N-terminal CBMx of *Ri*Xyn10A displays approximately 7-fold lower affinity for X6 than the average affinity measured for the enzyme variant lacking this module (Table 2). Nonetheless, CBMx is selective for arabinoxylan and contributes to the overall affinity of the enzyme (Table 1). Low-affinity CBMs may potentiate multivalent cooperative substrate binding, with less reduction of k_{cat} (that is, promote relatively high k_{cat}/k_{off}[33]), compared to higher-affinity counterparts that evoke higher energetic penalty during substrate displacement. The extended binding mediated by the CBMs seems to confer an advantage in the capture and prolonged contact of the enzyme with xylan. Deletion of the CBMs (*Ri*Xyn10A-cata) caused a substantial decrease in the apparent affinity towards WAX and BGX and deviation from Michaelis–Menten kinetics (Table 1 and Supplementary Fig. 4b–g). These findings are consistent with the importance of CBMs in catalysis under substrate limitations, whereas similar turnover rates were obtained by the catalytic module and the full-length *Ri*Xyn10A at high (9 mg ml^{-1}) substrate concentrations. Multiplicity and variability of CBMs seem to be a signature of extracellular enzymes from butyrate-producing Firmicutes[34,35]. By contrast, *Bacteriodes* members possess simpler xylanases with an inserted tandem CBM4 repeat within the catalytic module[23]. However, xylan capture by *Bacteriodes* is additionally orchestrated by moderate-affinity ($K_d \approx 60\,\mu M$) binding proteins[22].

R. intestinalis was able to compete with *B. ovatus* for xylans during the log-phase (Fig. 5e–g). Notably, *R. intestinalis* seemed to outcompete *B. ovatus* after propagation of the co-culture (in the late log phase) in fresh medium for two additional passages, underscoring the competitiveness of the xylan utilization machinery of this Firmicute (Supplementary Fig. 9c). *R. intestinalis* association to insoluble xylans, including wheat bran was reported, whereas *Bacteriodes* spp. were enriched in solubilized xylan fractions[18,36]. The extended binding by *Ri*Xyn10A may have an important role in the association to insoluble substrates. Indeed, the expression of this enzyme seemed similarly high in the mono- and mixed xylan cultures with *B. ovatus* (Supplementary Fig. 9e), in contrast to the reported downregulation of hydrolases by the taxonomically realted *E. rectale* during co-growth with *Bacteriodes thetaiotamicron* on a fibre-rich diet in mice[37].

The gene encoding the binding protein (*Ri*XBP) of the ABC transporter that confers xylo-oligosaccharide uptake in *R. intestinalis* was the most upregulated in the xylan transcriptomes, attesting the crucial role of oligosaccharide capture and transport in the densely populated gut ecological niche. The affinity and size preference of *Ri*XBP were found to be very different from the corresponding protein from *Bifidobacterium*[29], which prefers shorter xylo-oligosaccharides with a different side-chain decoration pattern. Importantly, striking differences in binding affinities and preference are observed when *Ri*XBP is compared to the SusD-like xylan-binding counterpart from *Bacteriodes*. Indeed, both SusD-like proteins from *B. ovatus*, which mediate the capture and internalization of xylo-oligosaccharides ≥X6 by SusC TonB-dependent permeases, displayed no measurable binding to X4 and X5 (ref. [22]), the preferred ligands of *Ri*XBP (Table 3). These differential transport protein preferences are likely to be instrumental in establishing competitive uptake profiles to select oligosaccharides of specific sizes and decorations for each taxon. This is supported by the dominance of *R. intestinalis* when the co-culture with *B. ovatus* was grown on X4 (Fig. 5h).

Our study highlights the molecular apparatus that *R. intestinalis*, as a model *Clostridium* group XIVa Firmicute, has evolved to compete for abundant dietary glycans with other dominant commensal bacteria. Strikingly, complex enzymes with multiple ancillary modules mediate multivalent substrate capture and breakdown. Highly expressed ABC transporters mediate efficient capture and uptake of xylan oligosaccharides with a different preference than the corresponding transport systems of currently known competing taxa. These findings suggest that the differentiation of glycan capture and uptake preferences may contribute to co-growth on dietary fibres by establishing competitive windows for distinct fibre breakdown oligomers by different taxa. Our study sets the stage for further investigations to evaluate the generality of this hypothesis.

Methods

Chemicals. All chemicals were of analytical grade. BGX, beechwood glucuronoxylan, corncob xylo-oligosaccharides and xylose were from Carl Roth (Karlsruhe, Germany). CBX was a kind gift from M. Yadav (US Department of Agriculture, Agricultural Research Service). Soluble wheat arabinoxylan (low viscosity: 10 centiStokes (cSt)) (WAX), insoluble wheat arabinoxylan (high viscosity: 48 cSt) (InWAX), xylobiose through to xylohexaose (X2–X6), arabinoxylotriose (AX3), arabinoxylotetraose (AX4) and mannohexaose (Man6) were from Megazyme. D-Glucuronic acid was from Sigma-Aldrich. L-Arabinose was from VWR International. Xylo-oligosaccharides Longlive 95 P (XOS) were from Shandong Longlive Bio-technology. AcBGX, acetylated aspen glucuronoxylan and acetylated spruce galactoglucomannan were prepared with steam explosion as previously described[38]. Cellulose acetate was a kind gift from A. Deutschle (University of Hamburg, Germany). Acetylated chitin-oligosaccharides were prepared as previously described[39].

Growth experiments and RNA-seq transcriptional analysis. *R. intestinalis* DSM 14610 was grown in a Whitley DG250 Anaerobic Workstation (Don Whitley Scientific) in YCFA medium[14] supplemented with autoclaved-sterilized 0.5% (w/v) carbohydrates. Cultures (5 ml) were grown in triplicates, and optical denisty at 600 nm (OD_{600}) and pH (for insoluble substrates) were measured to assess bacterial growth until the stationary phase was reached. Growth rates were calculated from the exponential growth phase.

For the RNA-seq analysis, total RNA was extracted at mid- to late-log phase (OD_{600} = 0.5–0.7) from biological triplicate cultures (10 ml) grown in YCFA

supplemented with 0.5% (w/v) glucose, xylose, WAX or BGX. Cells were harvested (4,000*g* for 5 min at room temperature) and the pellets were frozen at −80 °C until RNA extraction. The RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol after enzymatic lysis followed by mechanical disruption of the cells. A DNase treatment was included to ensure removal of DNA. The purity and quantity of the extracted RNA were assessed by an Agilent 2100 Bioanalyzer (Agilent Technologies). Removal of ribosomal RNA and library construction for RNA-seq were performed using the ScriptSeq Complete Kit (Epicentre). High-throughput sequencing was performed in a single lane in paired-end reads on an Illumina Hiseq 4000 platform at Beijing Genomics Institute (BGI). In total, 400 million paired-end reads were obtained and the read quality was assessed by FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The R1 reads were chosen for downstream analysis. Adaptor trimming and de-multiplexing were performed using custom python scripts (based on the Biopython SeqIO module[40]) and the FASTX-Toolkit v0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were further trimmed with fastx_trimmer and subsequently filtered with fastq_quality_filter with minimum quality score of 30 (-q 30), where 95% of base pairs (bp) meet the minimum quality score (-p 95). The resulting reads were kept if they were longer than 20 bp (-m 20). The *R. intestinalis* L1-82 reference genome and genome annotations are based on assembly GCA_000156535.1_ASM15653v1, obtained from the NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/Roseburia_intestinalis/). Reads were mapped to the reference genome using Tophat2 (refs [41,42]), and gene counts were determined with HTseq[43]. Differential gene expression was performed using DeSeq2 in R[44].

Xylanase activity measurements on whole cells. Cell-associated xylanase activity was determined by growing *R. intestinalis* cells in 800 µl YCFA containing 0.5% (w/v) xylo-oligosaccharides, WAX, BGX or glucose for 15 h. Cells were harvested (4,000*g* for 5 min at room temperature), resuspended in PBS to $OD_{600}=0.3$ and xylanase activity was assayed using the 3,5-dinitrosalicylic acid (DNS) assay as described below. To determine the effect of high ionic strength on the localization of xylanase activity, *R. intestinalis* cells were grown in 6 ml YCFA containing 0.5% (w/v) BGX for 15 h. Subsequently, the culture was divided into two 3 ml aliquots and harvested as described above. Cell pellets were resuspended in 300 µl PBS with or without 1.5 M NaCl. The suspensions were spun down and both pellets and supernatants (wash fractions) were collected. Cell pellets were washed with excess PBS and resuspended in 300 µl PBS. The xylanase activity of cells and wash fractions was assayed using the DNS assay.

Expression and purification of *R. intestinalis* proteins mediating xylan utilization. Open reading frames of the proteins without signal peptide, as predicted by SignalP v.3.0 (http://www.cbs.dtu.dk/services/SignalP-3.0), were amplified from *R. intestinalis* DSM 14610 genomic DNA using specific primers (Supplementary Table 13). Amplicons were cloned into the EcoRI and NcoI restriction sites of a pETM-11 (a kind gift from G. Stier, EMBL, Center for Biochemistry, Heidelberg, Germany)[45] or the XhoI and NcoI restriction site of a pET28a(+) (Novagen) using In-Fusion cloning (Takara) to express proteins as fusions with either cleavable N-terminal His_6 tags or C-terminal His_6 tags, respectively. Standard protocols were used for recombinant protein expression and purification using His-affinity and size-exclusion chromatography.

Enzymatic activity assays. Enzymatic assays were carried out in a 50 mM HEPES 0.005% (v/v) Triton X-100, pH 7.0 standard assay buffer unless otherwise stated. Hydrolysis kinetics of full-length or truncated xylanases (10–200 µM) were assayed towards 1–9 mg ml^{-1} of BGX, WAX or InWAX (900 µl at 37 °C for 12 min). Initial hydrolysis rates were determined by removing 200 µl aliquots every 3 min and quenching the reaction in 300 µl DNS reagent[46]. Next, samples were incubated for 15 min at 90 °C followed by absorbance at 540 nm (A_{540}) measurement in 96-well microtitre plates. Xylose was used as a standard (0–2.5 mM). Xylanase activity was assayed for *R. intestinalis* cells washed with PBS ± 1.5 M NaCl, and wash fractions, as above with the following modification: 180 µl of 1% (w/v) BGX was incubated with 20 µl cell suspension or wash fraction for 4 h.

Hydrolysis kinetics of α-glucuronidase were analysed on 1–9 mg ml^{-1} beechwood glucuronoxylan or a hydrolysate thereof (prepared by incubation with 4 mM *Ri*Xyn10A xylanase for 15 h at 37 °C followed by heat inactivation). The initial rates of MeGlcA release were measured using a coupled enzymatic assay (Megazyme). Reactions (770 µl) were incubated for 2 min at 37 °C with 10–180 nM enzyme with intermittent removal of 175 µl aliquots every 15 s into 125 µl 1 M Tris pH 10 to quench the reaction. This was followed by mixing 270 µl of the stopped reaction with 45 µl of the NAD^+ and uronate dehydrogenase reagents. Conversion of NAD^+ to NADH was measured at A_{340}. Glucuronic acid was used as standard (0–500 µM).

Hydrolysis kinetics of *Ri*Xyl8 and *Ri*Xyl3A were determined towards X2 through to X6 (0.5–12 mM) in McIlvaine buffer pH 6.8 (10 mM citric acid and 20 mM sodium phosphate) as described previously[47,48]. Reactions (350 µl) were incubated for 12 min at 37 °C with 36–78 nM *Ri*Xyl3A or 2.4 nM *Ri*Xyn8. Aliquots of 50 µl were removed every 2 min and stopped in 250 µl p-bromoaniline (2% w/v) in glacial acetic acid with thiourea (4% w/v). The stopped reactions were incubated in darkness for 10 min at 70 °C, followed by incubation at 37 °C for 1 h before measuring A_{520}. The concentration of released pentoses was determined using a xylose standard (0–5 mM)[49].

α-L-Arabinofuranosidase activity for *Ri*Abf43A was assayed in McIlvaine buffer pH 6.8 (10 mM citric acid and 20 mM sodium phosphate) using a coupled enzymatic L-arabinose/D-galactose assay (Megazyme) towards WAX (1–24 mg ml^{-1}). Reactions (75 µl) were incubated for 12 min at 37 °C with 0.4–1.7 µM enzyme. Aliquots of 15 µl were removed every 2 min, and the enzyme was inactivated (10 min at 90 °C) and thereafter 10 µl of this solution were mixed with 10 µl of the provided NAD^+, 20 µl of provided assay buffer and 2 µl galactose mutaotase/β-galactose dehydrogenase mix. The formation of NADH was measured as above. Arabinose was used as standard (0–5 mM).

The acetyl esterase-specific activity of *Ri*AXE was determined in 250 µl reactions containing para-nitrophenyl-acetate (4 mM) and 0.14 µM enzyme. A_{405} was measured every 60 s for 10 min at 37 °C in a microtitre plate reader and pNP (0–1 mM) was used as standard. The specific activity was determined in units (U) per mg, where a U is defined as the amount of enzyme that produces 1 µmol of pNP min^{-1}.

Kinetic parameters were calculated by fitting the Michaelis–Menten equation to the initial rate data using GraphPad Prism 7. The catalytic efficiency k_{cat}/K_m, determined from the slope of the normalized initial rate (V_0/[E]) in the Michaelis–Menten plot, is reported when saturation was not attained. All experiments were performed in triplicates.

Action patterns of individual and mixtures of xylanolytic enzymes. Hydrolysis of xylan and xylo-oligosachharides was performed at 37 °C for 15 h in the standard assay buffer used above. Oligosaccharide hydrolysates, used to assay the sequential action of the debranching xylanolytic enzymes, were generated using *Ri*Xyn10A, which was separated by ultrafiltration (3 kDa cut-off) before the addition of debranching enzymes. The hydrolysis profiles were analysed as detailed below. To verify the mode of reducing-end attack of *Ri*Xyl8, 30 mg XOS in standard assay buffer were reduced by $NaBH_4$ (1 M in 100 µM NaOH). A total of 200 µl of the $NaBH_4$ was added dropwise to 800 µl of the xylo-oligosaccharides solution, which was kept on ice. As a control, 100 µM NaOH was added to an 800 µl xylo-oligossaccharides solution. The mixture was incubated for 1 h at room temperature, then quenched by 400 µl 1 M acetic acid and diluted 10× in assay buffer.

Matrix-assisted laser desorption/ionization. Oligosaccharides were analysed with an Ultraflex matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF)/TOF instrument (Bruker Daltonics). The samples were applied with 2,5-dihydroxybenzoic acid (DHB) as matrix to a MTP 384 ground steel target plate (Bruker Daltonics). All spectra were obtained in positive reflection mode and processed using Bruker flexAnalysis 3.3.

Thin layer chromatography and HPAEC-PAD. Aliquots of 1 µl of enzymatic reactions were spotted on a silica gel 60 F254 plate (Merck). The chromatography was performed in a butanol:acetic acid:water (2:1:1 v/v) mobile phase. The plates were dried at 50 °C and carbohydrate hydrolysis products were visualized by spraying with a 5-methylresorcinol:ethanol:sulfuric acid (2:80:10 % v/v) developer and tarred briefly at 350 °C until bands appeared. Release of xylo-oligossaccharides and monosaccharides was analysed by HPAEC-PAD on an ICS-3000 (Dionex) using a 3 × 250 mm CarboPac PA1 column, a 3 × 50 mm guard column and 10 µl injections. Xylo-oligosaccharide and standards were eluted with mobile phase of constant 0.1 mM NaOH (flow rate: 0.35 ml min^{-1}) and a two-step linear gradient of sodium acetate: 0–25 min of 0–75 mM and 25–30 min of 75–400 mM. Monosaccharides and standards (0.1 mg ml^{-1}) of galactose, arabinose, glucose and xylose were eluted with 1 mM KOH for 35 min at 0.25 ml min^{-1}.

NMR spectroscopy. For the time-resolved NMR recordings: 4 mg AcBGX or acetylated spruce galactoglucomannan were dissolved in 500 µl 50 mM phosphate buffer pH 7.0 (99.9% D_2O). 2.5 µL of *Ri*AXE to a final concentration of 64 nM was added. The recorded spectrum is a pseudo-two dimensional (2D)-type experiment recording a 1D proton NMR spectrum every 5 min with in total 220 time points. The 1D proton spectrum was recorded with 24 scans using a 30° flip angle and a relaxation delay of 1 s (total recording time of 73 s). For enzyme treatment, 2.5 µl of *Ri*Xyn10A and *Ri*Agu115A were added to the AcBGX sample to 167 nM and 13 nM, respectively, and the sample was incubated at 37 °C for 24 h prior to *Ri*AXE addition. All homonuclear and heteronuclear NMR experiments were recorded on a BRUKER AVIIIHD 800 MHz (Bruker BioSpin) equipped with 5 mm with cryogenic CP-TCI, and all acquisitions were done at 37 °C. For the chemical shift assignment of AcBGX, the following spectra were recorded: 1D proton, 2D double-quantum-filtered correlation spectroscopy (DQF-COSY), 2D total correlation spectroscopy (TOCSY), 2D ^{13}C-heteronuclear single-quantum coherence (HSQC), 2D ^{13}C-heteronuclear 2 bond correlation (H2BC), 2D ^{13}C-HSQC-[^{1}H,^{1}H]TOCSY and 2D heteronuclear multiple bond correlation (HMBC). The acetate signal to 1.903 p.p.m. (pH 7.0 at 37 °C, in relation to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS)[50]) was used as chemical shift reference for protons, whereas ^{13}C chemical shifts were referenced indirectly to acetate, based on the absolute

frequency ratios[51]. The spectra were recorded, processed and analysed using TopSpin 3.5 software (Bruker BioSpin).

SPR. Xylo-oligosaccharide binding to *Ri*Xyn10A, *Ri*Xyn10AΔCBMx and *Ri*Xyn10A-CBMx was analysed using SPR on a BIAcore T100 (GE Healthcare). Immobilization of the proteins on a CM5 chips was performed using a random amine coupling kit (GE Healthcare) according to the manufacturer's protocol with 50–150 μg ml^{-1} protein in 10 mM sodium acetate pH 3.6–4.2, to a density of 1,362, 10,531 and 4,041 response units (RU) for *Ri*Xyn10AΔCBMx, *Ri*Xyn10A and *Ri*Xyn10A-CBMx, respectively. The analysis comprised 90 s of association and 240 s of dissociation at 30 μl min^{-1}. Sensograms were recorded at 25 °C in 20 mM phosphate/citrate buffer (pH 6.5), 150 mM NaCl and 0.005% (v/v) P20 (GE Healthcare). Solutions were filtered prior to analysis (0.22 μm). Experiments were performed in duplicates with seven concentrations in the range: 156 μM–10 mM for X3, 75 μM–4 mM for X4, X6 and Man6 and 62.5 μM–4 mM for X5. Data analysis was carried out using the Biacore T100 evaluation software, and K_d were determined by fitting a one-binding site model to the steady-state sensograms.

ITC. Titrations were performed using a Microcal ITC200 calorimeter (GE Healthcare) at 25 °C with *Ri*XBP (0.1 mM) or *Ri*Xyn10AΔCBMx (0.25 mM) in the sample cell and xylo-oligosaccharides (2.2–5 mM) in 10 mM sodium phosphate pH 6.5 in the syringe. An initial injection of 0.5 μl was followed by 19 × 2 μl injections separated by 120 s. The data were corrected for the heat of dilution, determined from buffer titration and a non-linear single-binding model was fitted to the normalized integrated binding isotherms using the MicroCal Origin software v7.0 to determine the binding thermodynamics.

Affinity electrophoresis. Binding of CBMx to WAX (0–0.1% w/v) or BGX (0–1.0% w/v) was assessed by affinity electrophoresis[52] in 10% native polyacrylamide gels (70 V for 3 h at 4 °C) using purified recombinant *Ri*Xyn10A-CBMx (3.0 μg) and β-lactoglobulin (1.5 μg) as a negative control. The relative mobility (r) was calculated as the migration of *Ri*Xyn10A-CBMx relative to migration of the dye front. A linear regression of the 1/r versus xylan concentration allowed the determination of K_d as the intercept of this *x* axis.

Western blot and immunofluorescence microscopy. Custom antibodies against the recombinant for the two xylanases *Ri*Xyn10A, *Ri*Xyn10B and the transport protein *Ri*XBP were raised in rats and rabbit, respectively (Eurogentec). The specificity of the antibodies was tested by western blots using a standard protocol. The membranes were blocked for 1 h in 1% (w/v) BSA in TBST buffer (Tris-buffered saline, 0.1% (v/v Tween 20)) and incubated for 2 h with the anti-sera (500× dilution in TBST-buffer). Subsequently, the membranes were washed three times in TBST buffer and incubated for 2 h with 6,000× diluted secondary polyclonal goat anti-rabbit IgG-AP antibodies coupled to alkaline phosphatase (AP) (Dako) and rabbit anti-rat IgG-AP (Sigma). After three washes, the proteins were visualized by exposure to Sigma-Fast BCIP/NBT reagent (Sigma).

R. intestinalis cells were grown in 6 ml YCFA containing 0.5% (w/v) WAX to $OD_{600} \approx 0.8$, harvested (4,000*g* for 5 min at room temperature) and washed twice in PBS. The cells were resuspended in 3 ml 4% (w/v) paraformaldehyde in PBS and fixed by incubation on ice for 15 min. Thereafter, the cells were washed twice in PBS and resuspended in 2 ml PBS. 50 μl of cell suspension were added to glass slides coated with poly-L-lysine, cells were blocked for 1 h in blocking buffer (1% (w/v) milk powder in PBS) and washed twice in PBS. For labelling, the cells were incubated with 50 μl anti-sera diluted 50× in blocking buffer for 2 h, washed twice in PBS and incubated for 1 h with 50 μl goat anti-rat IgG Alexa-Flour 555 or goat anti-rabbit IgG Alexa-Flour 488 (Thermo Scientific). Secondary antibodies were diluted 500× PBS. Finally, cells were washed twice in PBS, one drop of ProLong Gold antifade (Thermo Scientific) was applied and the cells were secured with a cover slide. Fluorescence was visualized using Zeiss Axioplan 2 microscope equipped with a CoolSNAP cf color camera and a Zeiss Plan-Neofluar × 100/1.3 NA, oil immersion objective.

Co-culture competition assay. *B. ovatus* DSM 1896 and *R. intestinalis* DSM 14610 were grown anaerobically in 20 ml YCFA supplemented with 0.5% (w/v) glucose to late-log phase, and an approximately equal number of cells (estimated by OD_{600}) were inoculated into CFA medium (YCFA lacking the yeast extract to minimize *B. ovatus* growth on yeast extract[53]) containing 0.5% (w/v) WAX, BGX, InWAX or X4. The co-cultures were grown in triplicates and samples (2 ml) were taken during growth. In the propagation experiment, the co-culture was passaged into fresh media after 9 h of growth (start $OD_{600} = 0.01$), then grown for 12 h and passaged again into fresh media and grown for 12 h. Genomic DNA was extracted from samples using DNAClean Microbial DNA isolation kit (Qiagen). Relative bacterial abundance was estimated by qPCR. The extracted DNA was diluted to 0.5 ng μl^{-1} and amplified in technical triplicates using strain-specific primers (Supplementary Table 14). The amplification mix contained 2 μl DNA, 5.5 μl LightCycler 480 SYBR Green I Master mix (Roche), 0.22 μl of each primer (10 pmol μl^{-1}) and 3 μl sterile water. Amplification conditions were 1 cycle of 95 °C for 5 min, 45 cycles of 95 °C for 10 s, 60 °C for 15 s and 72 °C for 45 s using a LightCycler 480 II (Roche). Relative bacterial concentrations were estimated by comparing gene copy numbers calculated using standard curves prepared with the respective reference DNA. Western blot was performed as described above but with cell cultures instead of purified proteins.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Code availability. Adaptor trimming and demultiplexing was performed using custom python scripts (based on the Biopython SeqIO module[40]) and the FASTX-Toolkitv0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). The fastx_barcode_splitter script (from FASTX-toolkit) was modified to demultiplex large fastq files provided from BGI. The modified version of this script is available upon request.

Data availability. The proteins characterized in this study are available from the NCBI with the following accession numbers: EEV01588.1 (ROSINTL182_06494), EEU99940.1 (ROSINTL182_08193), EEU99941.1 (ROSINTL182_08194), EEU99942.1 (ROSINTL182_08195), EEU99943.1 (ROSINTL182_08196), EEU99894.1 (ROSINTL182_08199) and EEU99897.1 (ROSINTL182_08202). The authors declare that the data supporting the findings of this study are available within the paper and the supplementary information or from the corresponding author on request. The fastx_barcode_splitter script (from the FASTX-toolkit) was modified to demultiplex large fastq files provided from the BGI. The modified version of this script is available upon request.

Received: 25 September 2017; Accepted: 19 February 2018;
Published online: 2 April 2018

References

1. Nicholson, J. K. et al. Host–gut microbiota metabolic interactions. *Science* **336**, 1262–1267 (2012).
2. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016).
3. Marchesi, J. R. et al. The gut microbiota and host health: a new clinical frontier. *Gut* **65**, 330–339 (2016).
4. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2013).
5. Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339–1353 (2016).
6. Cockburn, D. W. & Koropatkin, N. M. Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *J. Mol. Biol.* **428**, 3230–3252 (2016).
7. Xu, S. et al. Butyrate induces apoptosis by activating PDC and inhibiting complex I through SIRT3 inactivation. *Signal Transduct. Target. Ther.* **2**, e16035 (2017).
8. Donohoe, D. R. et al. The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol. Cell* **48**, 612–626 (2012).
9. Furusawa, Y. et al. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013).
10. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
11. Takahashi, K. et al. Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. *Digestion* **93**, 59–65 (2016).
12. Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
13. Vrieze, A. et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143**, 913–916 (2012).
14. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615–1620 (2002).
15. Louis, P. & Flint, H. J. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol. Lett.* **294**, 1–8 (2009).
16. Van den Abbeele, P. et al. Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J.* **7**, 949–961 (2013).
17. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11**, 497–504 (2013).
18. Mirande, C. et al. Dietary fibre degradation and fermentation by two xylanolytic bacteria *Bacteroides xylanisolvens* XB1AT and *Roseburia intestinalis* XB6B4 from the human intestine. *J. Appl. Microbiol.* **109**, 451–460 (2010).

19. Chassard, C., Goumy, V., Leclerc, M., Del'homme, C. & Bernalier-Donadille, A. Characterization of the xylan-degrading microbial community from human faeces. *FEMS Microbiol. Ecol.* **61**, 121–131 (2007).
20. Sheridan, P. O. et al Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genomics* **2**, e000043 (2016).
21. Selvendran, R. R. Chemistry of plant cell walls and dietary fibre. *Scand. J. Gastroenterol.* **5521**, 33–41 (1987).
22. Rogowski, A. et al. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
23. Zhang, M. et al. Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl Acad. Sci. USA* **111**, E3708–E3717 (2014).
24. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
25. Kelly, G. et al. Structure of the cell-adhesion fragment of intimin from enteropathogenic *Escherichia coli*. *Nat. Struct. Mol. Biol.* **6**, 313–318 (1999).
26. Ebbes, M. et al. Fold and function of the InlB B-repeat. *J. Biol. Chem.* **286**, 15496–15506 (2011).
27. Karlsson, E. N. et al. The modular xylanase Xyn10A from *Rhodothermus marinus* is cell-attached, and its C-terminal domain has several putative homologues among cell-attached proteins within the phylum Bacteroidetes. *FEMS Microbiol. Lett.* **241**, 233–242 (2004).
28. Ejby, M. et al. An ATP binding cassette transporter mediates the uptake of α-(1,6)-linked dietary oligosaccharides in *Bifidobacterium* and correlates with competitive growth on these substrates. *J. Biol. Chem.* **291**, 20220–20231 (2016).
29. Ejby, M. et al. Structural basis for arabinoxylo-oligosaccharide capture by the probiotic *Bifidobacterium animalis* subsp. lactis Bl-04. *Mol. Microbiol.* **90**, 1100–1112 (2013).
30. Honda, Y. & Kitaoka, M. A family 8 glycoside hydrolase from *Bacillus halodurans* C-125 (BH2105) is a reducing end xylose-releasing exo-oligoxylanase. *J. Biol. Chem.* **279**, 55097–55103 (2004).
31. Sun, Z. et al. A novel three-component system-based regulatory model for D-xylose sensing and transport in *Clostridium beijerinckii*. *Mol. Microbiol.* **95**, 576–589 (2015).
32. Duncan, S. H. et al. Wheat bran promotes enrichment within the human colonic microbiota of butyrate-producing bacteria that release ferulic acid. *Environ. Microbiol.* **18**, 2214–2225 (2016).
33. Morrill, J. et al. The GH5 1,4-β-mannanase from *Bifidobacterium animalis* subsp. lactis Bl-04 possesses a low-affinity mannan-binding module and highlights the diversity of mannanolytic enzymes. *BMC Biochem.* **16**, 26 (2015).
34. Cockburn, D. W. et al. Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*. *Mol. Microbiol.* **95**, 209–230 (2015).
35. Ze, X. et al. Unique organization of extracellular amylases into amylosomes in the resistant starch-utilizing human colonic firmicutes bacterium *Ruminococcus bromii*. *mBio* **6**, e01058-15 (2015).
36. De Paepe, K., Kerckhof, F.-M., Verspreet, J., Courtin, C. M. & Van de Wiele, T. Inter-individual differences determine the outcome of wheat bran colonization by the human gut microbiome. *Environ. Microbiol.* **19**, 3251–3267 (2017).
37. Mahowald, M. A. et al. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl Acad. Sci. USA* **106**, 5859–5864 (2009).
38. Biely, P. et al. Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. *Biochim. Biophys. Acta* **1830**, 5075–5086 (2013).
39. Sørbotten, A., Horn, S. J., Eijsink, V. G. H. & Vårum, K. M. Degradation of chitosans with chitinase B from *Serratia marcescens*. *FEBS J.* **272**, 538–549 (2005).
40. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
41. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
43. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
44. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
45. Dümmler, A., Lawrence, A.-M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.* **4**, 34 (2005).
46. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428 (1959).
47. Roe, J. H. & Rice, E. W. A photometric method for the determination of free pentoses in animal tissue. *J. Biol. Chem.* **173**, 507–512 (1948).
48. Deschatelets, L. & Yu, E. K. C. A simple pentose assay for biomass conversion studies. *Appl. Microbiol. Biotechnol.* **24**, 379–385 (1986).
49. Dilokpimol, A. et al. Enzymatic synthesis of β-xylosyl-oligosaccharides by transxylosylation using two β-xylosidases of glycoside hydrolase family 3 from *Aspergillus nidulans* FGSC A4. *Carbohydr. Res.* **346**, 421–429 (2011).
50. Govind, V., Young, K. & Maudsley, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **13**, 129–153 (2000).
51. Zhang, H., Neal, S. & Wishart, D. S. RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **25**, 173–195 (2003).
52. Takeo, K. Affinity electrophoresis: principles and applications. *Electrophoresis* **5**, 187–195 (1984).
53. Scott, K. P., Martin, J. C., Duncan, S. H. & Flint, H. J. Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro. *FEMS Microbiol. Ecol.* **87**, 30–40 (2014).
54. Anand, S., Kaur, H. & Mande, S. S. Comparative in silico analysis of butyrate production pathways in gut commensals and pathogens. *Front. Microbiol.* **7**, 1945 (2016).

Acknowledgements

We thank B. Henrissat, architecture et fonction des macromolécules biologiques, CNRS, Aix-Marseille University and the curator of CAZy, for his advice and discussions on the assignment of the novel CBMx and the esterase. We also thank M. Yadav, US Department of Agriculture, Agricultural Research Service, for the kind gift of cornbran xylan, and BioCHOS AS for providing the chitooligo (CHOS) sample. M. Due, T. Holm Madsen and C. Aaarup Christensen are thanked for their technical help in cloning recombinant proteins and the performance of binding experiments. We also wish to thank A. Schultz, H. Juel Martens and M. Hansen, PLEN, University of Copenhagen, for the use of the confocal laser scanning microscopy in the initial microscopy experiments. This project was funded by a Graduate School DTU Scholarship, Lyngby, Denmark. Additional fundings were from the Danish Research Council for Independent Research, Natural Sciences (DFF, FNU) by a Research Project 2 grant (grant ID: 4002-00297B), a BIONÆR project (grant number: 244259) and the Norwegian NMR Platform, NNP (F.L.A.) from the Research Council of Norway (226244). Carlsberg Foundation is acknowledged for an ITC instrument grant (2011-01-0598).

Author contributions

Growth analysis was performed by M.L.L. Transcriptomic analysis was done by M.L.L., C.W. and D.A.E. Enzyme characterization was performed by M.L.L., M.E., S.S.P., F.L.A. and B.W. qPCR was done by M.L.L. and M.I.B. Microscopy was perfromed by M.L.L. and C.S. Experiments were designed by M.L.L. and M.A.H. The manuscript written by M.L.L. and M.A.H. with contributions from T.R.L., B.W. and F.L.A. Figures were prepared by M.L.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-018-0132-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.A.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature microbiology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41564-018-0132-8

In the format provided by the authors and unedited.

Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut

Maria Louise Leth[1], Morten Ejby[1], Christopher Workman[1], David Adrian Ewald[1], Signe Schultz Pedersen[1], Claus Sternberg[1], Martin Iain Bahl[2], Tine Rask Licht[2], Finn Lillelund Aachmann[3], Bjørge Westereng[4] and Maher Abou Hachem[1]*

[1]Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark. [2]National Food Institute, Technical University of Denmark, Lyngby, Denmark. [3]NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway. [4]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway. *e-mail: maha@bio.dtu.dk

Supplementary Table 2. Modular organization of GH10 xylanases from human gut Firmicutes and Bacteroidetes.

Phylum	Family	Strain	Accession number	Length (AA)	CBMs
Firmicutes	Lachnospiraceae	*Roseburia intestinalis* L1-82	ROSINTL182_06494	1356	X, 22, 9, 9
			ROSINTL182_6338-9	601	
		Roseburia intestinalis XB6B4	CBL13458.1	1356	X, 22, 9, 9
		Roseburia intestinalis M50/1	n.a.	1356	X, 22, 9, 9
		Roseburia faecis M72	CRL32809.1	1380	X, 22, 9, 9
		Eubacterium rectale T1-815	CRL34489.1	1028	X, 9, 9
		Butyrivibrio fibrisolvens 16/4	CBK74925.1	1153	9
			CBK75021.1	690	13, 2
		Hungatella hathewayi	CUO52114.1	421	
		Ruminococcus gnavus	WP_064787180.1	394	
	Ruminococcaceae	*Ruminococcus champanellensis* 18P13	CBL16579.1	633	22
			CBL17682.1	1268	22, 22, 6
		Ruminococcus callidus ATCC 27760	ERJ94429.1	1158	22, 22, 9
			ERJ87773.1	630	22
			ERJ97032.1	382	22
Bacteroidetes	Bacteroidaceae	*Bacteroides ovatus*	EDO13863.1	372	
			EDO10007.1[1]	376	
			EDO14247.1	573	
			EDO10010.1[1]	740	4, 4
			EDO14052.1	584	
			EDO10798.1	750	
		Bacteroides intestinalis DSM 17393	EDV05054.1	782	4, 4
			EDV05072.1[2]	746	4, 4
			EDV03684.1	738	
			EDV05059.1	910	
			EDV07678.1	725	
			EDV07007.1[2]	899	
		Bacteroides xylanisolvens XB1A	CBK67953.1[3]	754	4, 4
			CBH32823.1	378	

AA: amino acids, n.a.: GH10 is present, but not assigned in the genome,

1. Rogowski, A. *et al.* Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6,** 7481 (2015).
2. Zhang, M. *et al.* Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111,** E3708-E3717 (2014)
3. Despres, J. *et al.* Xylan degradation by the human gut Bacteroides xylanisolvens XB1AT involves two distinct gene clusters that are linked at the transcriptional level. *BMC Genomics* **17,** 326 (2016).

Supplementary Table 3. Thermodynamic parameters and dissociation constant for *Ri*Xyn10A-CBMx determined by ITC.

Ligand	K_D (µM)	N_0	ΔH (kcal/mol)	TΔS(kcal/mol)	ΔG (kcal/mol)
X6	413 ± 125	0.74 ± 0.04	-19.9 ± 1.2	-15.3	-4.6

Data are from one experiment and binding parameters are reported with the error of the fit to the binding isotherm.

Supplementary Table 4. Homologs of CBMx identified in genomes of taxonomically related taxa to *R. intestinalis*.

Strain	Accession number	Query cover	E-value	Identity
Roseburia intestinalis XB6B4	CBL13458.1	100%	4e-85	100%
Eubacterium rectale_T1815	CRL34489.1	89%	5e-36	55%
Butyrivibrio sp. LC3010	WP_026509692.1	92%	1e-07	36%
Roseburia faecis M72	CRL32809.1	93%	9e-12	36%
Bacterium enrichment culture clone MC3F	AFU34339.1	86%	3e-07	30%
Lachnoclostridium phytofermentans ISDg	ABX41884.1	84%	5e-07	26%
Clostridium sp. KNHs205	WP_033165005.1	88%	1e-06	28%
Butyrivibrio sp. INlla14	SCX91715.1	63%	2e-06	32%
Lachnospiraceae bacterium YSD2013	SCX14282.1	73%	1e-05	34%
Butyrivibrio sp. ob235	SEK63083.1	76%	2e-04	30%
Butyrivibrio sp. VCD2006	WP_026526370.1	72%	3e-04	27%

Supplementary Table 5. Kinetic parameters of *Ri*Agu115A.

Substrate	K_M (mg mL^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M (mL mg^{-1} s^{-1})
BeGX	n.d.	n.d.	2
BeGX + *Ri*Xyn10A	12 ± 3	395 ± 34	33

n.d.: Low affinity and lack of curvature of the Michaelis Menten plots precluded reliable determination of kinetic parameters. Catalytic efficiencies are from the slope of the initial rates versus substrate concentration. Data are means of a triplicate with standard deviations.

Supplementary Table 6. Kinetics of *Ri*Abf43A.

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
AX4	0.8 ± 0.1	20 ± 1	25
	K_M (mg mL^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M (mL mg s^{-1})
WAX	6.3 ± 0.4	12 ± 0	1.9

Data are means of a triplicate with standard deviations.

Supplementary Table 7. Kinetics *Ri*Xyl3A.

Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} mM^{-1})
X2	2.7 ± 0.4	57 ± 3	21
X3	3.4 ± 0.3	60 ± 2	18
X4	2.4 ± 0.4	32 ± 2	13
X5	2.6 ± 0.5	36 ± 1	14
X6	2.1 ± 0.2	30 ± 1	15

Data are means of a triplicate with standard deviations.

Supplementary Table 8. Kinetics *Ri*Xyl8.

Substrate	K_M (mg/mL)	k_{cat} (S^{-1})	k_{cat}/K_M ($s^{-1}mM^{-1}$)
X3	4.8 ± 1.0	1208 ± 124	251.7
X4	5.1 ± 1.5	892 ± 131	174.9

Data are means of a triplicate with standard deviations.

Supplementary Table 9. Deacetylation activity of *Ri*AXE on acetylated xylans and aryl acetate.

Substrate	Enzyme(s)	*V* (μM s^{-1})	*V*/[E] (s^{-1})
AcBGX	*Ri*AXE	2.5	39.1
	*Ri*AXE+*Ri*Xyn10A	3.2	50
	*Ri*AXE+*Ri*Xyn10A+*Ri*Agu115A	2.8	43.8
AcSpruce mannan	*Ri*AXE	0.2	3.1
pNP-acetate	*Ri*AXE	4.7[a]± 0.1	n.d.
Autolysis		0.07	n.d.

V: rate, *V*/[E]: normalized rate by enzyme concentration estimated from NMR experiments. [a]The activity on paranitrophenyl acetate (pNP-acetate) is expressed in U mg^{-1}.

Supplementary Table 10. Assignment of chemical shifts for xylan deacetylation by *Ri*AXE.

Structural unit	Assignment						
	H-1; C-1	H-2; C-2	H-3; C-3	H-4; C-4	H-5; C-5	H-6; C-6	Ac-H; C
X	4.42; 105.4	3.19; 75.4	3.53; 76.4	3.78; 79,2	n.d	n.d	-
C2	4.68; 102.6	4.69; 76.1	3.79; 74.2	3.86; 78.9	n.d	n.d	2.10; 23.1 /2.16; 23.1
C3	4.47; 104.3	3.37; 75.4	4.89; 79.9	3.78; 79.1	n.d	n.d	2.17;23.2
C23	4.81; 102.2	4.81; 74.2	5.17; 74.1	4.05; 77.9	n.d	n.d	(2) 2.10; 22.9/ (2)2.12; 23.0
C3MeGlcA	4.57; 104.2	3.48; 73.6	4.98; 78.1	3.94;78.1	n.d	n.d	2.15; 23.3
MeGlcA	5.17; 96.6	3.56; 74.4	3.53; 73.3	n.d	n.d	n.d	-
α	5.18; 94.8	3.56;74.2	3.53;73.7	n.d	n.d	n.d	-
β	4.56; 99.3	3.25;76.7	3.52;77.9	3.72;79.7	n.d	n.d	-

Supplementary Table 11. Esterase activity for *Ri*AXE measured using MALDI-TOF.

	AcBGX	AcAspen xylan	AcSpruce mannan	Cellulose mono acetate	AcChitin	InWAX
***Ri*AXE**	++	++	+	+	-	-
***Ri*AXE + *Ri*Agu115A**	+++	n.d.	n.d.	n.d.	n.d.	n.d.

+++: complete deacetylation, ++:almost complete acetylation (1 ≥ acetyl/oligosaccharide),

+: minor deactylation (1-2 acetyl/oligosaccharide), -: no deacetylation). Experiments performed twice.

Supplementary Table 12. Xylan hydrolysis kinetics of *Ri*Xyn10B.

Substrate	K_M (mg mL^{-1})	k_{cat} (s^{-1})	k_{cat}/K_M (mL mg^{-1} s^{-1})
BGX	n.d.	n.d.	9.8
WAX	4.4 ± 0.8	413 ± 32	94
InWAX	n.d.	n.d.	2.3

n.d.: Low affinity and lack of curvature of the Michaelis Menten plots precluded reliable determination of the kinetic parameters and the catalytic efficiencies are determined from the slope of the initial rate data versus substrate concentration. Data are reported as means of triplicates with standard deviations

Supplementary Table 13. Cloning and mutagenesis primers[a,b].

Gene	Accession number	Name	Orientation	Sequence (5' -> 3')
ROSINTL182_ 06494 (AA27-1356)	EEV01588.1	*Ri*Xyn10A	Forward	**TTTCAGGGCGCCATG**GGGGTAAAAAAAGTTTTTACTGCAGAT
ROSINTL182_ 06494 (AA27-1356)	EEV01588.1	*Ri*Xyn10A	Reverse	**GACGGAGCTCGAATTTTA**CTACTTACTGATCTTTATCTTCTTTGCA
ROSINTL182_ 06494 (AA156-1356)	EEV01588.1	*Ri*Xyn10AΔCBMx	Forward	**TTTCAGGGCGCCATG**GCAGGAGCAGGCGATGCA
ROSINTL182_ 06494 (AA156-1356)	EEV01588.1	*Ri*Xyn10AΔCBMx	Reverse	**GACGGAGCTCGAATTTTA**CTACTTACTGATCTTTATCTTCTTTGCA
ROSINTL182_ 06494 (AA349-754)	EEV01588.1	*Ri*Xyn10A-cata	Forward	**TTTCAGGGCGCCATG**TCTATTGAGAAGGACATCCCGGA
ROSINTL182_ 06494 (AA349-754)	EEV01588.1	*Ri*Xyn10A-cata	Reverse	**GACGGAGCTCGAATTTTA**GGATGCATCTACATACGCCCA
ROSINTL182_ 06494 (AA27-165)	EEV01588.1	*Ri*Xyn10A-CBMx	Forward	**TTTCAGGGCGCCATG**GGGGTAAAAAAAGTTTTTACTGCAGAT
ROSINTL182_ 06494 (AA27-165)	EEV01588.1	*Ri*Xyn10A-CBMx	Reverse	**GACGGAGCTCGAATTTTA**ATCCCCCAATTTTGCA
ROSINTL182_ 08193	EEU99940.1	*Ri*Abf43A	Forward	**AGGAGATATACCATG**AGTATAGCAAAGAATCCGGTTC
ROSINTL182_ 08193	EEU99940.1	*Ri*Abf43A	Reverse	**GGTGGTGGTGCTCGA**AACCCGGTATTCCCTCATA
ROSINTL182_ 08194	EEU99941.1	*RI*AXE	Forward	**AGGAGATATACCATG**AGTGGACCTGTGGCA
ROSINTL182_ 08194	EEU99941.1	*RI*AXE	Reverse	**GGTGGTGGTGCTCGA**ATTCCACATAGCCAAAACCAA
ROSINTL182_ 08195	EEU99942.1	*Ri*Agu115A	Forward	**TTTCAGGGCGCCATG**GAAGCAATTTTGGTAAAGGATC
ROSINTL182_ 08195	EEU99942.1	*Ri*Agu115A	Reverse	**GACGGAGCTCGAATTTTA**TCATCTGTTCGTCCTCCTT
ROSINTL182_ 08196	EEU99943.1	*Ri*Xyl8	Forward	**AGGAGATATACCATG**AAAAGAGGAGCGTTTGAGA
ROSINTL182_ 08196	EEU99943.1	*Ri*Xyl8	Reverse	**GGTGGTGGTGCTCGA**AATAAATTCTATAATTGCCGCTCAG
ROSINTL182_ 08199	EEU99894.1	*Ri*XBP	Forward	**TTTCAGGGCGCCATG**GGAAACAAAGCAGCCG
ROSINTL182_ 08199	EEU99894.1	*Ri*XBP	Reverse	**GACGGAGCTCGAATTTTA**TTACTGATATTTTTTTGCTTCCTC
ROSINTL182_ 08202	EEU99897.1	*Ri*Xyl3A	Forward	**AGGAGATATACCATG**GAATTAAATCAGAATACAGAAAAACTG
ROSINTL182_ 08202	EEU99897.1	*Ri*Xyl3A	Reverse	**GGTGGTGGTGCTCGAA<u>TAA</u>**CATCAGACTTTCCACTGTTT
ROSINTL182_06338/ ROSINTL182_06339	EEV01752.1/ EEV01731.1	*Ri*Xyn10B	Forward	**TTTCAGGGCGCCATG**GCTGGGCAGGAAAATG
ROSINTL182_06338/ ROSINTL182_06339	EEV01752.1/ EEV01731.1	*Ri*Xyn10B	Reverse	**GACGGAGCTCGAATTTTA**CTATTTATCAGAATGAAATAAATTTTCAA

[a]Bold nucleotides indicate the sequences annealing to the vector.
[b]Underlined nucleotides indicate the changed codon and italics indicate the changed bases.

Supplementary Table 14. qPCR primers use.

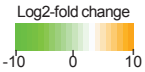
Target bacteria	Orientation	Sequence (5' -> 3')	Reference
Roseburia spp.	Forward	TACTGCATTGGAAACTGTCG	[1]
Roseburia spp	Reverse	CGGCACCGAAGAGCAAT	[1]
Bacteroides spp.	Forward	CGATGGATAGGGGTTCTGAGAGGA	[2]
Bacteroides spp.	Reverse	GCTGGCACGGAGTTAGCCGA	[2]
Universal primer	Forward	ACTCCTACGGGAGGCAGCAGT	[3]
Universal primer	Reverse	GTATTACCGCGGCTGCTGGCAC	[3]

1. Larsen, N. *et al.* Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* **5,** e9085 (2010).
2. Bergström, A. *et al.* Introducing GUt Low-Density Array (GULDA)-a validated approach for qPCR-based intestinal microbial community analysis. *FEMS Microbiol. Lett.***337,** 38–47 (2012).
3. Walter, J. *et al.* Detection and identification of gastrointestinal *Lactobacillus* species by using denaturing gradient gel electrophoresis and species-specific PCR primers. *Appl. Environ. Microbiol.***66,**297–303 (2000).

Supplementary Figures

a

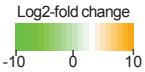
Locus ID	Log2-fold change			SP	Protein	Annotation
	X1/Glc	WAX/Glc	BGX/Glc			
05034	-0.84	6.48	5.54	Yes	SBP	ABC transporter, solute-binding protein
05035	-0.79	5.99	5.44	No	PP	ABC transporter, permease protein
05036	-0.80	6.04	5.26	No	PP	ABC transporter, permease protein
05037+05106	-0.33	6.44	5.95	Yes		NHL repeat protein
05107	-1.48	5.98	5.53	No		Hypothetical protein
05108	-0.42	5.91	5.48	No		Hypothetical protein
05109	-0.61	5.87	4.92	No	PP	ABC transporter, permease protein
05110	-0.63	5.11	3.92	No	PP	ABC transporter, permease protein
05111+05112	0.34	5.73	5.07	Yes	SBP	ABC transporter, solute-binding protein
05113	-0.77	4.95	4.78	No		Hypothetical protein
05114	-0.95	4.09	4.08	No	GH115	Xylan α-1,2-glucuronidase
05115	-0.73	4.35	4.17	No	CE	Putitativ esterase



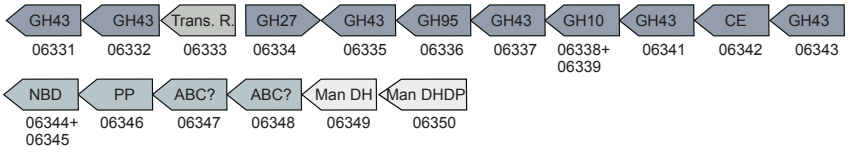
b

c

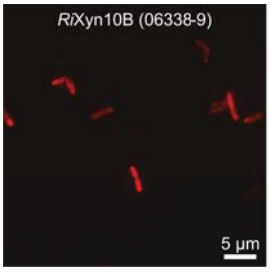
Locus ID	Log2-fold change			SP	Protein	Annotation
	X1/Glc	WAX/Glc	BGX/Glc			
06331	-0.63	5.26	4.76	No	GH43	β-xylosidase/α-L-arabinofuranosidase
06332	-1.60	5.46	4.74	No	GH43	β-xylosidase/α-L-arabinofuranosidase
06333	-0.70	0.44	2.24	No	AraC	Transcriptional regulator
06334	-1.22	1.03	2.44	No	GH27	α-galactosidase/β-L-arabinopyranosidase
06335	-0.69	3.58	3.44	No	GH43	β-xylosidase/α-L-arabinofuranosidase
06336	-1.04	3.61	3.21	No	GH95	α-L-galactosidase/α-L-fucosidase
06337	-0.57	3.45	4.08	No	GH43	β-xylosidase/α-L-arabinofuranosidase
06338+06339	-0.99	3.47	3.62	Yes	*Ri*Xyn10B	Endo-1,4-β-xylanase
06341	-1.12	4.09	3.80	No	GH43	β-xylosidase/α-L-arabinofuranosidase
06342	-1.32	4.28	4.07	No	CE1	Esterase
06343	-1.23	4.49	4.21	Yes	GH43	β-xylosidase/α-L-arabinofuranosidase
06344+06345	-1.57	4.49	4.68	No	ABC-NBD	ABC transporter, nucleotide binding domain
06346	-2.45	5.12	4.74	No	ABC-PP	ABC transporter, permerase protein
06347	-2.38	5.49	5.35	No		Hypothetical ABC transporter
06348	-1.62	4.52	4.04	No		Hypothetical ABC transporter
06349	-0.95	6.49	5.44	No	UxuA	Mannonate dehydratase
06350	-1.50	5.84	5.12	No	UxuB	Mannitol/D-arabinitol dehydrogenase domain protein



d

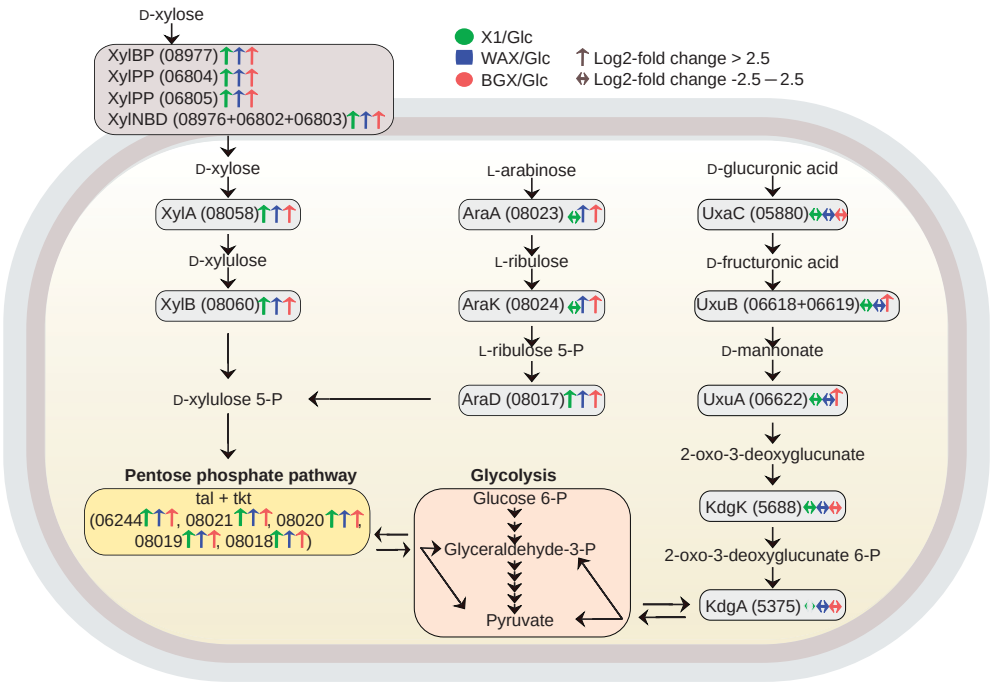


e



Supplementary Figure 1 *R. intestinalis* L1-82 unique xylan upregulated loci. (a) Upregulation of a putative xylan metabolism gene cluster unique for the *R. intestinalis* L1-82 strain on xylan. **(b)** Organization of genes in (a). **(c)** Second unique *R. intestinalis* L1-82 gene cluster upregulated on xylan. **(d)** Organization of putative xylan-metabolism genes upregulated in (c). **(e)** Fluorescence microscopy of *R. intestinalis* grown on xylan showing the extracellular localization of *Ri*Xyn10B. Experiments were performed three times and locus IDs ROSINTL182_xxxxx are abbreviated with the last numbers after the hyphen. Signal peptides (SP) were predicted using SignalP v.3.0. Genes residing between two contigs have two locus IDs.

a

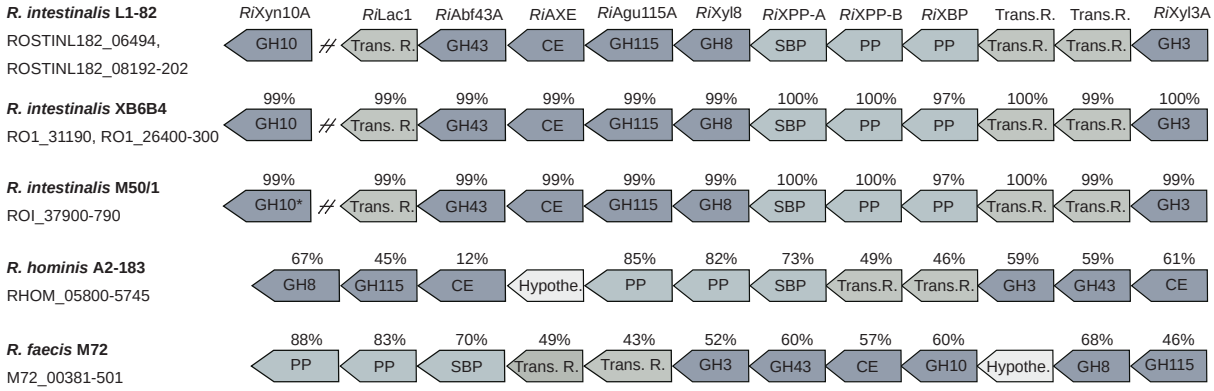


b

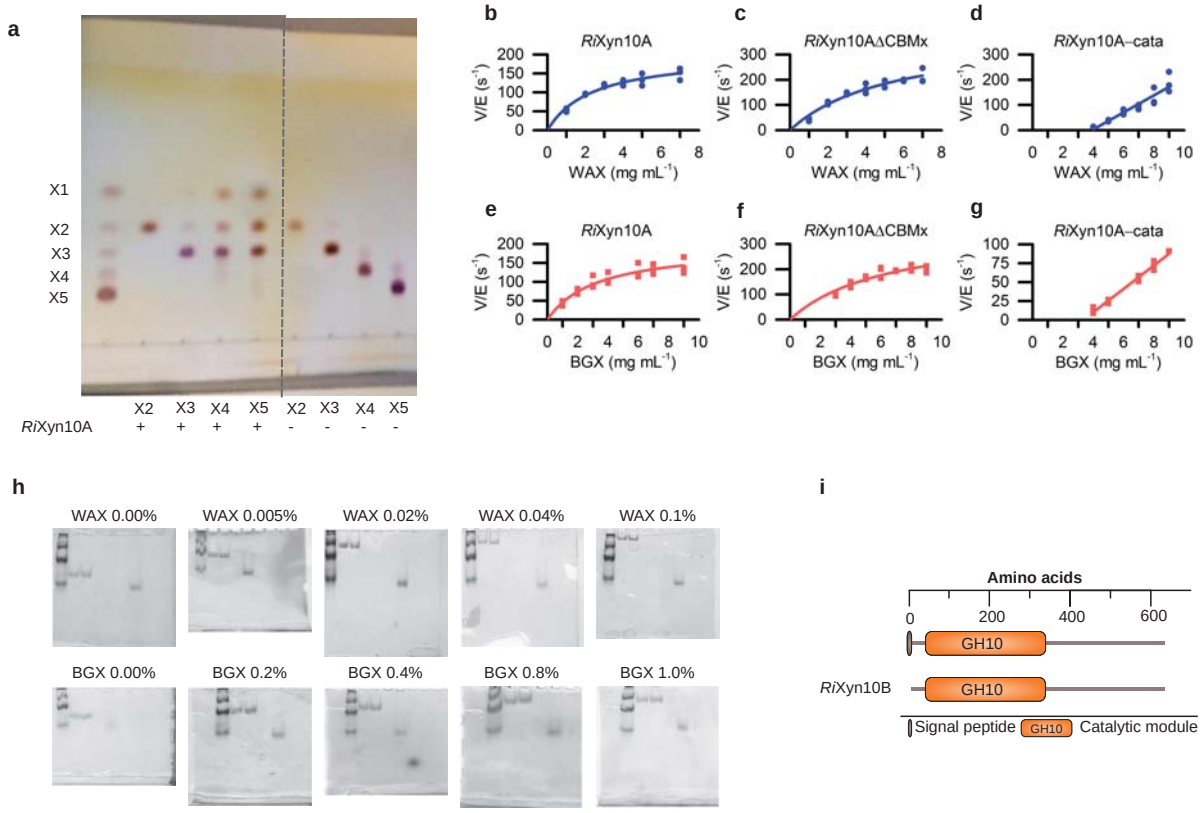
Locus ID	Log2-fold change			SP	Protein	Annotation
	X1/Glc	WAX/Glc	BGX/Glc			
08977	9.09	5.79	5.26	Yes	XylBP	ABC transporter, solute-binding protein
08976+06802+06803	8.98	6.15	5.65	No	XylNBD	ABC transporter, nucleotide binding domain
06804	8.73	5.56	4.43	No	XylPP	ABC transporter, permease protein
06805	8.42	5.50	4.75	No	XylPP	ABC transporter, permease protein
08058	5.51	4.34	4.16	No	XylA	Xylose isomerase
08060	5.05	4.28	3.77	No	XylB	Xylulokinase
08023	-0.66	5.89	3.11	No	AraA	L-arabinose isomerase
08024	-0.44	5.55	2.90	No	AraK	L-ribulokinase
08017	4.40	5.06	4.96	No	AraD	L-ribulose-5-phosphate 4-epimerase
05880	0.70	0.58	0.59	No	UxaC	Glucuronate isomerase
06618+06619	0.18	-0.24	5.77	No	UxuB	Mannonate oxidoreductase
06622	-1.06	1.11	4.06	No	UxuA	Mannonate dehydratase
05688	-0.18	-0.69	1.26	No	KdgK	2-keto-3-deoxy-D-gluconate kinase
05375	-0.61	-0.86	1.26	No	KdgA	2-dehydro-3-deoxy phosphogluconate aldolase
06244	6.80	7.15	5.93	No	tkt	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase
08021	5.29	5.72	4.85	No	tal	Fucose isomerase
08020	4.33	4.75	4.37	No	tkt	Transketolase
08019	4.06	4.44	3.76	No	tkt	Transketolase
08018	4.19	4.65	4.47	No	tkt	Transketolase

Log2-fold change
-10 0 10

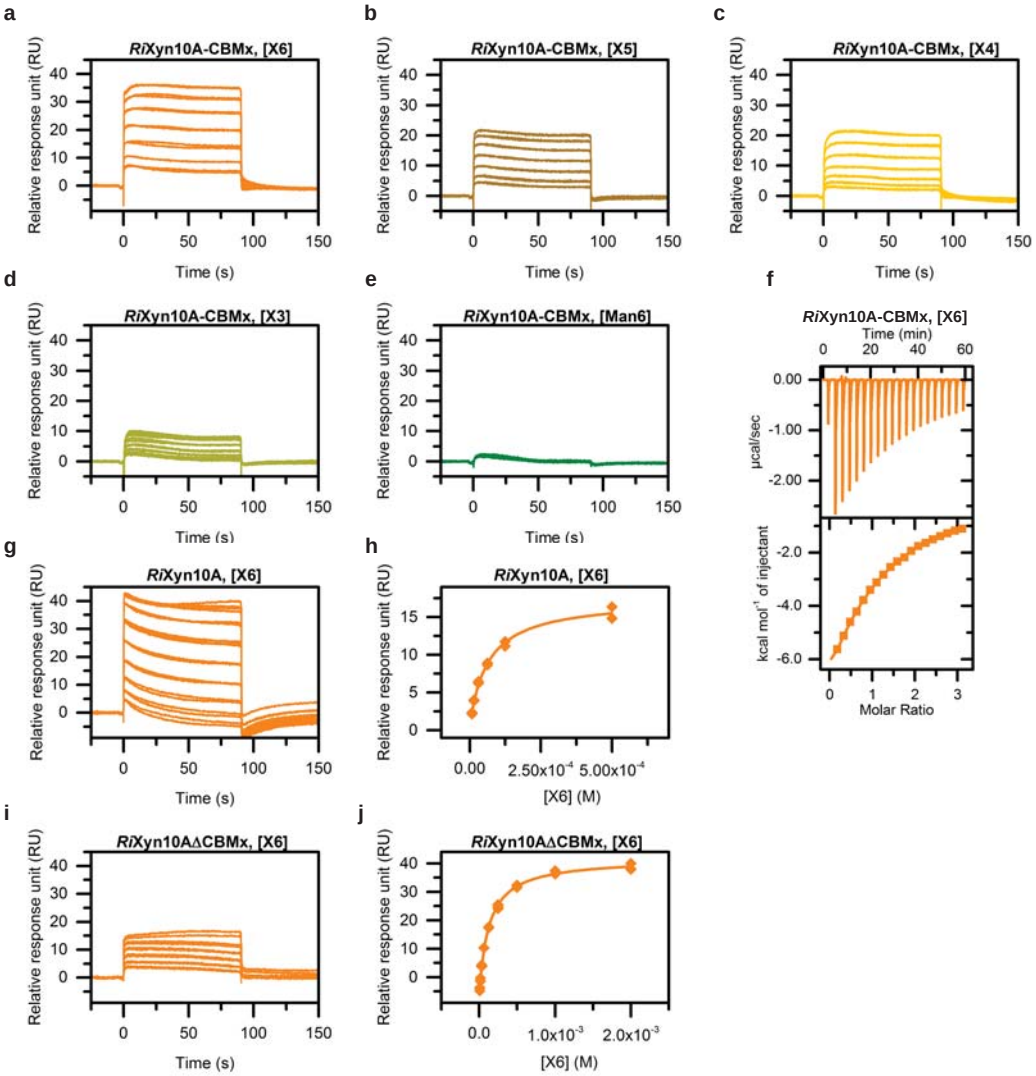
Supplementary Figure 2 *R. intestinalis* L1-82 xylose metabolism. (a) Proposed model for the metabolism of the monosaccharides xylose, arabinose and glucuronic acid in *R. intestinalis* L1-82 based on the RNA-seq data in Supplementary Table 1, and literature. **(b)** Upregulation of xylose import and metabolism genes in the model. The RNA-Seq heatmap depicts Log2-fold changes of genes expressed by cells grown on xylose (X1), wheat arabinoxylan (WAX) and birch glucuronoxylan (BGX) relative to glucose (Glc). Locus numbers ROSINTL182_xxxxx are abbreviated with the last numbers after the hyphen.



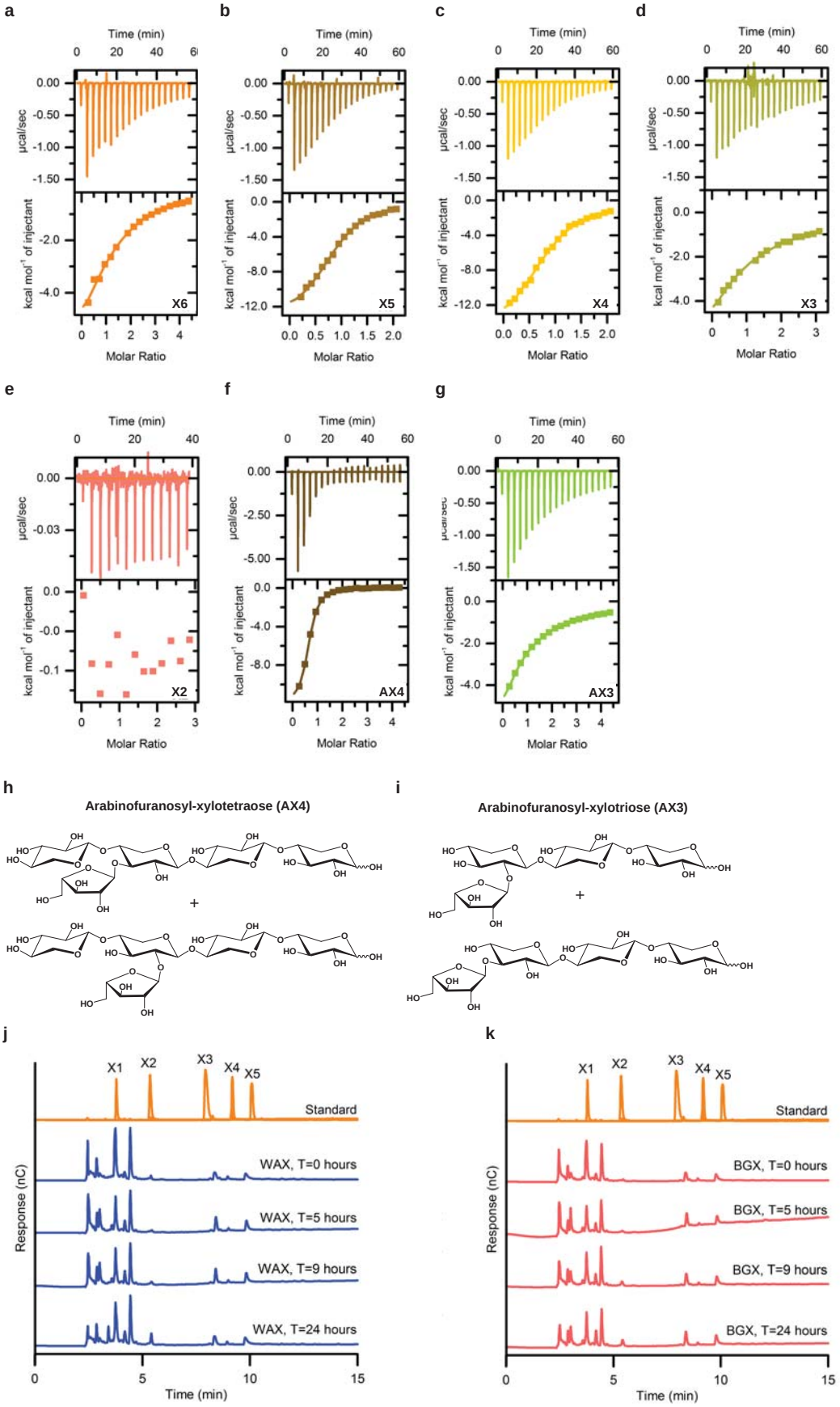
Supplementary Figure 3 Conservation of *R. intestinalis* core xylan utilization genes within the *Roseburia* genus. Genes are denoted according to their protein products; glycoside hydrolase (GH), carbohydrate esterase (CE), transcriptional regulators (Trans.R.), ABC transporter solute binding protein (SBP), ABC transporter permease protein (PP) and hypothetical proteins (Hypothe.). Sequence identities to *R. intestinalis* L1-82 genes are shown above the genes; Locus IDs for the genes are denoted under the respective strains. The asterisk indicates that the GH10 is not assigned in the genome.



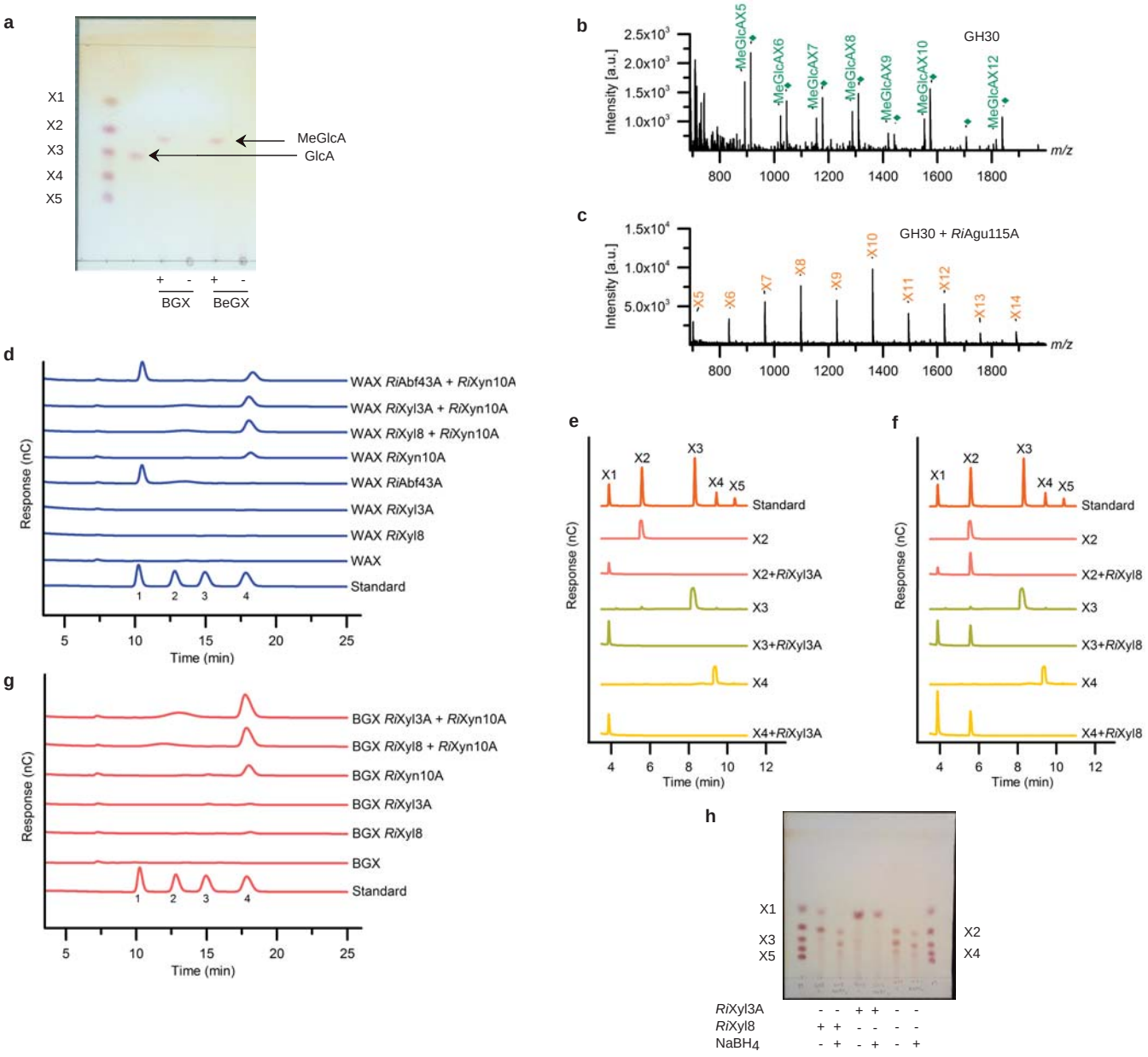
Supplementary Figure 4 Properties of the extracellular xylanases from *R. intestinalis* (a) Action patterns of *R.*Xyn10A on X2ĀX5 analyzed by TLC; +: reaction with enzyme, -: controls without enzyme. The dotted line indicates that lanes not relevant to the figure were spliced out for clarity. **(b-g)** Hydrolysis kinetics of *R.*Xyn10A, *R.*Xyn10A CBMx lacking the N-terminal module and *R.*Xyn10A-cata, the catalytic module on WAX, and BGX. **(h)** Binding of xylans to *R.*Xyn10-CBMx by affinity gel electrophoresis using native polyacrylamide gels with different concentrations of WAX (0.0-0.1% w/v) or BGX (0.0-1.0% w/v). No polysaccharides were added to the control. Lane 1+2; *R.*Xyn10A-CBMx (3.0 µg), Lane 3 -lactoglobulin (1.5 µg), M; marker. **(i)** Domain organization of the xylanase *R.*Xyn10B encoded by a locus upregulated on xylan and which is unique for the *R i.ntest.nal.s* L1-82 strain used in the present study (Supplementary Fig. 1c-d). The bottom cartoon represents the recombinant enzyme. Experiments in (a) and (h) are performed twice and in triplicates for (b-g).



Supplementary Figure 5 Binding of CBMx and *Ri*Xyn10A to xylo-oligosaccharides. (a-e) Reference and blank corrected sensograms depict binding of xylo-oligosaccharides (X3-X6) and mannohexaose (Man6) as negative control to CBMx (*Ri*Xyn10A-CBMx) using SPR analysis. **(f)** ITC analysis of CBMx binding to X6. **(g,i)** Reference and blank corrected SPR sensograms depicting the binding of X6 to *Ri*Xyn10A and *Ri*Xyn10AĀCBMx respectively. **(h,j)** One binding model fitted to the binding isotherms from the sensograms in (g,i). The experiments were in triplicates, except for the ITC run once.



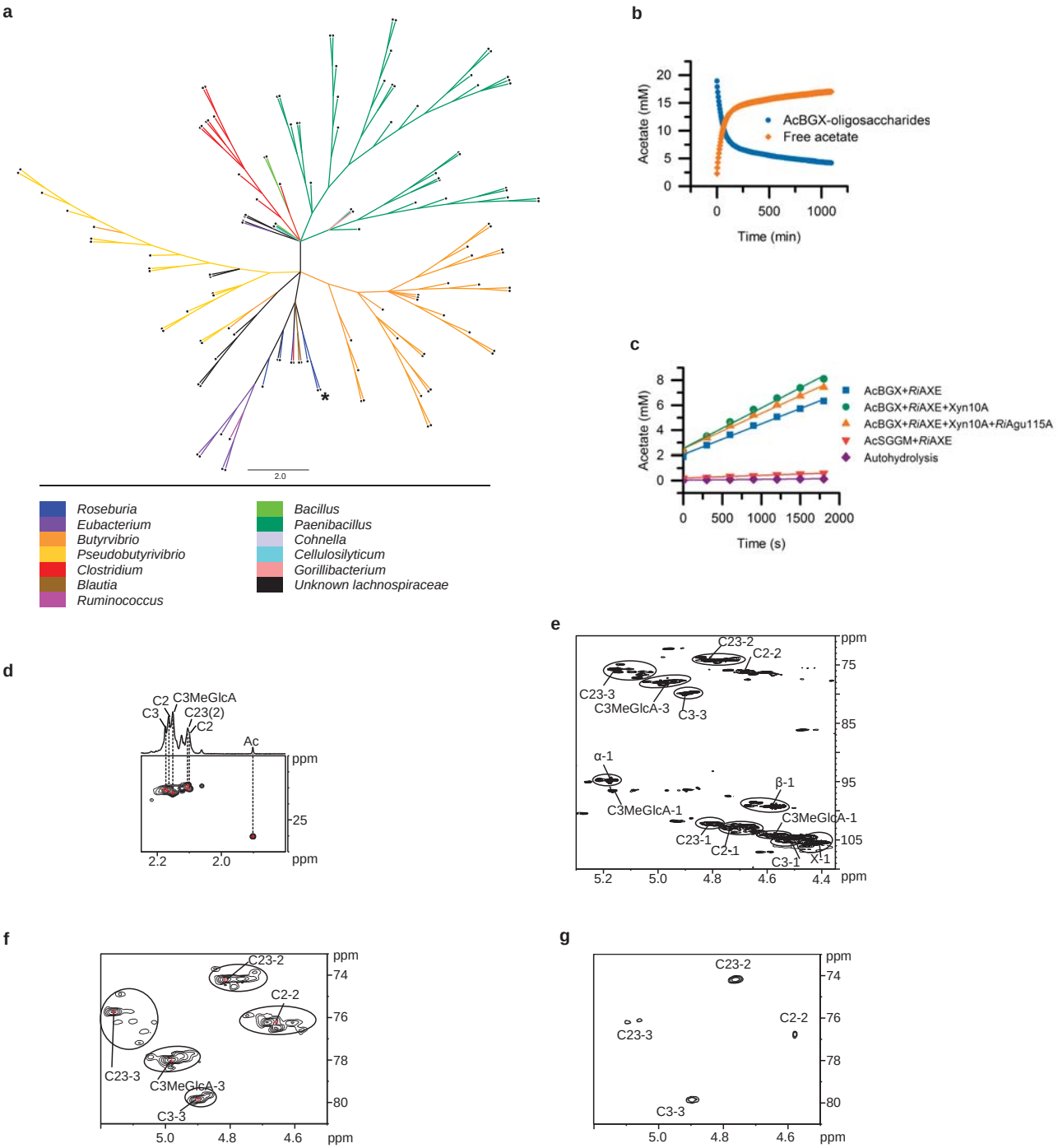
Supplementary Figure 6 Binding preference of *R.*XBP associated to the xylo-oligomer ABC transporter of *R i.ntest.nal.s*. (a-g) ITC analysis of *R.*XBP binding to linear and branched xylo-oligosaccharides. **(h,i)** Structures of the branched arabinosylated xylo-oligosaccharides AX4 and AX3, which are mixtures with arabinofuranosyl decoration either at the C2 or C3 of xylosyl units. **(j,k)** Time course HPAEC-PAD analysis of culture supernatants of *R i.ntest.nal.s* grown in YCFA with 0.5% WAX or BGX. The observed peaks between 0 and 5 minutes are likely unutilized medium components. Experiments in (a-g) are duplicates, and in (j,K) from a duplicate.



Supplementary Figure 7 Intracellular xylo-oligosaccharide degrading enzymes from *R. intestinalis* (a) TLC analysis of the release of 4-*O*-methylglucuronic acid (MeGlcA) from BGX and BeGX by *Ri*Agu115A. Glucuronic acid (GlcA) is used as standard. **(b,c)** Activity of *Ri*Agu115A on a GH30-hydrolyzed BeBGX monitored using MALDI-ToF MS; (b) is the GH30 control and (c) is the treatment with GH30 and *Ri*Agu115A. Activity indicates *Ri*Agu115A releases MeGlcA from the penultimate xyloxyl to the reducing end in xylo-oligosaccharides based on the GH30 strict specificity[1], whereas a GH10 generates xylo-oligosaccharides with a MeGlcA substitution at the non-reducing end[2]. This data shows that the *Ri*Agu115A is able to act on both internal and terminal non-reducing end substitutions on glucuronoxylan-derived xylo-oligosaccharides. Di-sodium adducts of MeGlcA decorated oligomers (diamonds) are colored as their corresponding single sodium adducts. **(d,g)** Monosaccharide hydrolysis products from enzymatic treatment of WAX and BGX with *Ri*Xyn10A, *Ri*Abf43A, *Ri*Xyl3A and *Ri*Xyl8 by HPAEC-PAD. Standards were 1; arabinose, 2; galactose, 3; glucose, 4; xylose. **(e,f)** *Ri*Xyl3A and *Ri*Xyl8 hydrolysis of xylo-oligosaccharides analyzed with HPAEC-PAD. **(h)** Ā-Xylosidase activity for *Ri*Xyl3A and *Ri*Xyl8 towards xylo-oligosaccharides (XOS) by TLC. The + and - indicate the presence and absence of the different components, respectively. Lack of activity on substrate reduced with NaBH4 (converts reducing end unit to its alditol) provided evidence that *Ri*Xyl8 acts on the reducing end as the alditol is not accommodated in the active site. Experiments are performed in duplicates.

1. St John, F. J., Hurlbert, J. C., Rice, J. D., Preston, J. F. & Pozharski, E. Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase. *J. Mol. Biol.* **407,** 92–109 (2011).

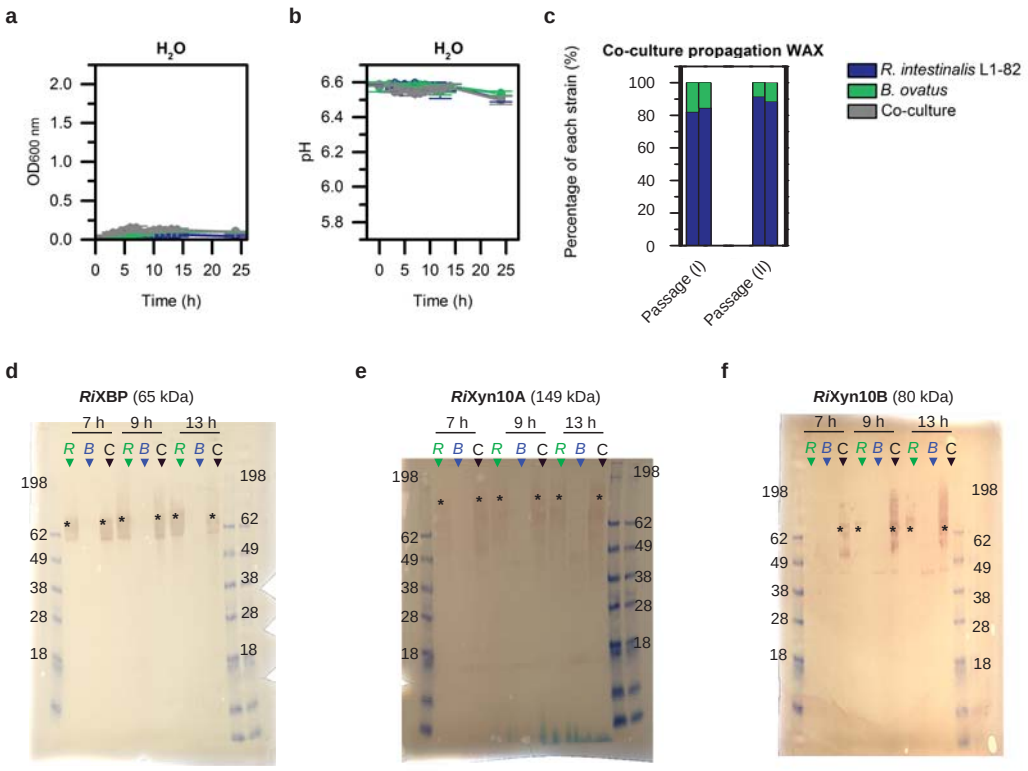
2. Dodd, D. & Cann, I. K. Enzymatic deconstruction of xylan for biofuel production. *Glob Chang. Biol Bioenergy* **1,** 2–17 (2009).



Supplementary Figure 8 Activity, specificity and taxonomic distribution of the novel xylan acetyl esterase *Ri*AXE. (a) Phylogenetic tree of *Ri*AXE and homologs identified by a BLASTP search against the non-redundant database. Sequences with coverage >86% and identity >42% were selected. All sequences were from Firmicutes members. The resulting 131 protein sequences were aligned using Muscle[1] and a phylogenetic tree constructed by the maximum likelihood algorithm in MEGA7[2]. Bootstraps were performed with 500 replicates. The phylogenetic tree was visualized using Figtree (http://tree.bio.ed.ac.uk/software/figtree). Asterisk indicates position of *Ri*AXE. **(b)** Time course deacetylation of AcBGX treated with *Ri*Xyn10A and *Ri*Agu115A by *Ri*AXE determined by NMR. **(c)** Rates of deacetylation by *Ri*AXE on AcBGX and AcSpruce mannan (AcSGGM) in D_2O, which may influence absolute reaction rates. **(d)** ^{13}C HSQC spectrum of *Ri*Xyn10A treated AcBGX showing the acetyl region and with the 1D proton projection. **(e)** same as (d) but showing the spectral region for anomeric and *O*-acetylated xylose signals. *Ri*Xyn10a treatment enhances signal-to-noise of resonances in the NMR spectra for the assignment and increases the total number of observable individual signals. **(f,g)** ^{13}C HSQC spectra for *O*-acetylated regions before (f) and after (g) deactylation by *Ri*AXE. Nearly complete deacetylation of AcBGX is reached during the time resolved NMR experiment. Chemical shifts of the most dominating signal for the monosaccharide residues mark by "+", peaks encircled by dotted lines indicate cluster of chemical shifts likely to belong to the same type of monosaccharide residue as dominating signal.

1. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).

2. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33,** msw054 (2016).



Supplementary Figure 9 Co-culture experiment with *R. intestinalis* and *B. ovatus*. (a-b) Growth curves for monoculture and co-cultures after growth of *R. intestinalis* and *B. ovatus* with water as controls instead of carbon source. **(c)** Relative abundance determined by qPCR in a propagation experiment with co-cultures on WAX. After 9 hours of growth, the co-culture was passaged into fresh media, passage (I) (start OD_{600} nm=0.01). This culture was grown for 12 hours and passaged into fresh media again (passage II). The western blots were carried out with **(d)** anti-*Ri*XBP, **(e)** anti-*Ri*Xyn10A, **(f)** anti-*Ri*Xyn10B. R: *R. intestinalis*, B: *B. ovatus*, C: co-culture of *R. intestinalis* and *B. ovatus*. Asterisk denotes the position of the band based on theoretical molecular mass. The molecular markers size is shown in kDa. Lower molecular mass signals than expected indicate proteolytic cleavage occurring particularly with the multi-modular *Ri*Xyn10A. Experiments are performed in biological triplicates in (a-c) and in duplicates in (d-f).

Chapter 3 Paper 2: ^{1}H, ^{13}C and ^{15}N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from *Roseburia intestinalis*

The present paper reports the backbone and side-chain assignment by NMR of the novel CBMx from a GH10, discovered in the paper "*Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut"* (Chapter 2). In addition, analysis of the secondary structure elements indicates three α-helices and ten β-sheets. The paper is prerequisite for the NMR study reported in chapter 4. The NMR experiments were performed by Finn Lillelund Aachmann and his group, while I contributed to the publication by cloning of the CBMx together with establishment of protein expression and purification conditions.

Biomolecular NMR Assignments
https://doi.org/10.1007/s12104-018-9850-3

ARTICLE

^{1}H, ^{13}C and ^{15}N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from *Roseburia intestinalis*

Eva Madland[1] · Yoshihito Kitaoku[2] · Gerd Inger Sætrom[1] · Maria Louise Leth[3] · Morten Ejby[3] · Maher Abou Hachem[3] · Finn Lillelund Aachmann[1]

Received: 19 July 2018 / Accepted: 17 September 2018

Abstract
The N-terminal domain (residues 28–165) from the glycoside hydrolase family 10 from *Roseburia intestinalis* (*Ri*CBMx), has been isotopically labeled and recombinantly expressed in *Escherichia coli*. Here we report ^{1}H, ^{13}C and ^{15}N NMR chemical shift assignments for this carbohydrate binding module (CBM).

Keywords Carbohydrate binding module (CBM) · Xylan binding module · Xylanase · *Roseburia intestinalis* · Gut microbiota

Biological context

Most of the dietary fibers in the human diet comes from the plant cell walls present in fruits and vegetables. Here we find complex polysaccharides such as, cellulose, pectin and xylan (Koropatkin et al. 2012). The latter has a β-1,4-linked xylose backbone with varying degrees of polymerization and side-chain substitution (Rennie and Scheller 2014). Xylan can be substituted with carbohydrate side-chains, e.g. arabinofuranosyl and glucuronosyl residues or acetyl groups. A variety of linkages to side-chain residues necessitate the deployment of different enzymes for xylan degradation.

Members of the human gut microbiota (HGM) are able to ferment xylan into short-chain fatty acids (SCFAs) e.g. butyrate, propionate and acetate. Butyrate is known to have a beneficial effect on the host's health by being an energy source for colonocytes as well as reducing the risk of colon cancer and enteric colitis (Donohoe et al. 2012; Morrison and Preston 2016; Xu et al. 2017). As the population of butyrate-producers are more abundant in healthy individuals, there is a particular interest in the role they play in the HGM (Sheridan et al. 2016). One of the key known butyrate-producers from the HGM is *Roseburia*, a common genus in the clostridial cluster XIVa within Firmicutes (Louis and Flint 2009). *Roseburia intestinalis* has shown xylanolytic activity, and is together with species from *Bacteroides*, one of the few taxa that can utilize xylan (Chassard et al. 2007; Mirande et al. 2010). The ability of *Bacteroides* to degrade xylan have been investigated in detail (Martens et al. 2011; Rogowski et al. 2015; Zhang et al. 2014), but insight into the strategy used by Firmicutes to harvest energy from xylan has until recently been lacking.

A recent study (Leth et al. 2018) showed that *R. intestinalis* is able to breakdown xylan through a modular cell-attached xylanase of glycoside hydrolase family 10 (*Ri*GH10A) that is conserved within the species. This enzyme is highly upregulated when *R. intestinalis* is grown on xylan and comprises of four xylan-binding modules: Two carbohydrate binding modules (CBMs) of family 9 (CBM9), one from family 22 (CBM22) and an N-terminal of a previously unknown family (CBMx). This representative of a new CBM family possesses low affinity for xylan, but displays selectivity for arabinoxylan, which makes it an interesting candidate for both structural and functional characterization studies by nuclear magnetic resonance (NMR).

✉ Finn Lillelund Aachmann
finn.l.aachmann@ntnu.no

[1] NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway

[2] Department of Advanced Bioscience, Kinki University, Nara, Japan

[3] Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark

Published online: 22 September 2018

Methods and experiments

Protein expression and purification

^{13}C, ^{15}N and ^{15}N samples were expressed in *Escherichia coli* BL21 (DE3). Pre-culture was grown in LB medium (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl) supplemented with 10 µL kanamycin (50 mg/mL) in a shaking incubator at 30 °C, 225 rpm overnight. From the pre-culture, 1% (v/v) was used to inoculate 450 mL M9 media (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl) supplemented with 500 µL kanamycin (50 mg/mL), 1 mL 1 M $MgSO_4$, 10 mL Trace Metal solution (0.1 g/L $ZnSO_4$, 0.8 g/L $MnSO_4$, 0.5 g/L $FeSO_4$, 0.1 g/L $CuSO_4$, 1 g/L $CaCl_2$), 5 mL Gibco™ MEM Vitamin Solution (100×), 10 mL ^{15}N Bioexpress Cell Growth Media (Cambridge Isotope Laboratories, Tewksbury, MA; USA), 2 g glucose (^{15}N label)/98% $^{13}C_6$-D-glucose (^{13}C, ^{15}N label) in 20 mL milliQ. After inoculation the medium was supplemented with 1 mL anti-foam and the cells were grown using Lex-24™ (Epiphyte) at 30 °C until $OD_{600} = 0.8$. The expression was induced with IPTG (isopropyl-1-thio-β-D-galactopyranoside) to a final concentration of 1 mM, and incubated with Lex-24™ (Epiphyte) at 16 °C over night. The cells were harvested by centrifugation (Sorvall) at 4 °C, 6000×*g*, 5 min. The pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 50 mM NaCl and 0.05% TritonX-100) and ½ tablet cOmplete™ ULTRA protease inhibitor (Roche) in 20 mL milliQ, and sonicated using a Branson Sonifier equipped with a microtip for 10 min. Isolation of the lysate was done by centrifugation (Eppendorf) at 4 °C, 15,000×*g*, for 2.5 h.

An Econo-Column® (Bio-Rad) containing 2 mL Ni–NTA Agarose (QIAGEN) was equilibrated with 20 column volumes WEB (50 mM Na_2HPO_4, 300 mM NaCl), pH 8.0. The lysate was incubated in the column for 45 min, and eluted with WEB with an increased amount of imidazole: 10 mM, 15 mM, 20 mM, 100 mM, 200 mM and 400 mM. Fractions containing CBMx were collected and purity confirmed with SDS–PAGE. To remove imidazole, the fractions were dialysed (MWCO 3.5 kDa) against 5 mM NaH_2PO_4, pH 5.5 over night.

To cleave the His-tag of the fraction containing CBMx, TEV-protease was added in 1/100 (w/w) and dialysed (MWCO 3.5 kDa) against 20 mM sodium phosphate, 1 mM DTT, 0.5 mM EDTA, pH 8.0. Purification of CBMx was done using a gravity flow column containing 2 mL cOmplete His-tag purification resin (Roche) equilibrated with the dialysis buffer. The dialyzed sample was applied to the column and the flow through was collected. The resin was washed with 1–5 column volumes of the same buffer and the sample was collected by combining this fraction with the flow through. Regeneration of the column was done with dialysis buffer containing 50 mM imidazole. An SDS–PAGE was run to confirm the separation and purity of the mature CBMx.

The CBMx containing fractions were concentrated and buffer exchanged into the NMR-buffer (50 mM sodium phosphate, pH 6.5. Samples for NMR was made with CBMx in NMR-buffer with D_2O added to a final ratio of 90% H_2O/10% D_2O) by centrifugation using Vivaspin® 6 protein spin concentrators (MWCO 5 kDa, Sartorius) at 10 °C, 7000×*g*.

NMR spectroscopy

All CBMx NMR samples were prepared in 50 mM sodium phosphate buffer and 10% D_2O, pH 6.5.

All spectra were recorded at 25 °C on a Bruker Ascend 800 MHz spectrometer Avance III HD Bruker Biospin equipped with a 5 mm Z-gradient CP-TCI (H/C/N) cryoprobe at the NV-NMR-Centre/Norwegian NMR Platform at NTNU Norwegian University of Science and Technology, Trondheim, Norway. ^1H shifts were referenced internally to HDO, while ^{13}C and ^{15}N chemical shifts were referenced indirectly to HDO, based on the absolute frequency ratios (Zhang et al. 2003). Backbone and side-chain assignments of CBMx were elucidated using ^{15}N-HSQC, ^{13}C-HSQC, HNCA, HNcoCA, HNCO, HNcaCO, HNCACB, HNcoCACB, HNHAHB, HNcoHAHB and HcCH-TOCSY. The NMR data were recorded and processed with TopSpin version 3.5 and the data was analyzed with CARA version 1.5 (Keller 2004). Secondary structure elements were evaluated using TALOS-N (https://spin.niddk.nih.gov/bax/software/TALOS-N/) (Shen and Bax 2013) and chemical shits of N, H^N, C^α, C^β, H^α, H^β and C'.

Assignment and data deposition

Here we report the backbone and side-chain assignment of CBMx. Figure 1 shows the ^{15}N-HSQC spectrum of CBMx together with the assigned resonances. The backbone assignment is essentially complete (N, HN, C^α, H^α and $C' > 98\%$). The mature protein contains two extra amino acids (Gly-Ala) at the N-terminus (for purification purpose) that were only partially assigned. Side-chain assignment is almost complete (H and C side-chains > 78%). The overall percentage of completion is affected by the missing assignment of exchangeable side-chain protons of Arg, Lys, Asn and Gln as well as aromatic residues. Chemical shift data have been deposited at the Biological Magnetic Resonance Data Bank (BMRB) under the accession number 27536.

Analysis of the secondary structure elements (Fig. 2) indicates three α-helices and ten β-sheets. A high degree of β-sheets is consistent with previously reported structures of

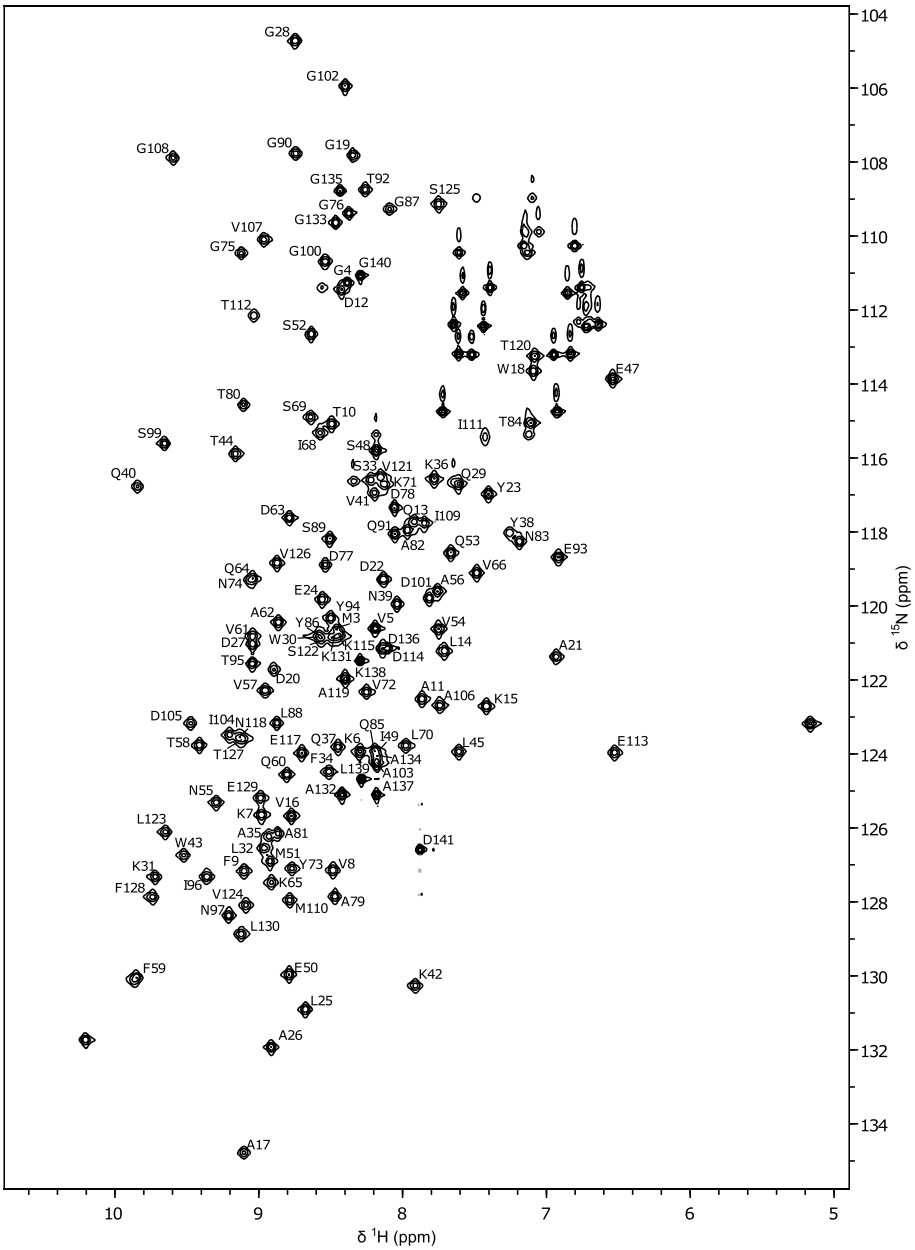


Fig. 1 ^{1}H, ^{15}N HSQC spectrum of ^{13}C, ^{15}N-labeled CBMx (1.7 mM) from the glycoside hydrolase family 10 xylanase from *Roseburia intestinalis* (*Ri*Xyn10A) in 50 mM sodium phosphate, pH 6.5 with D_2O added to a final ratio of 90% H_2O/10% D_2O. Residue number and type are indicated on the figure

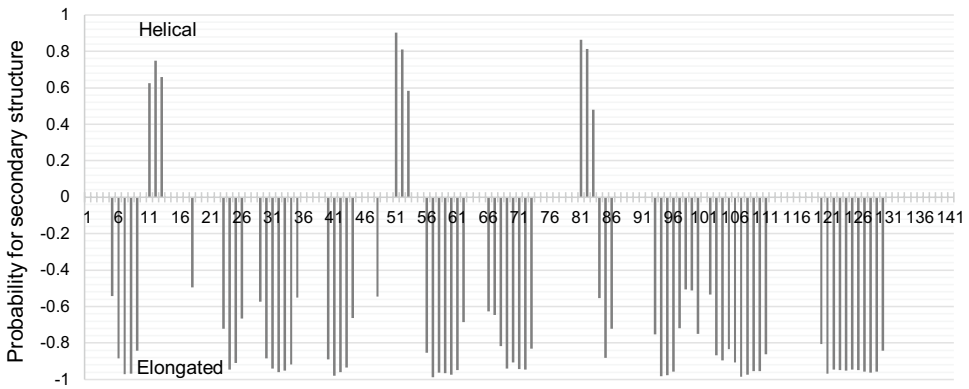


Fig. 2 Secondary structure probability of CBMx using TALOS-N. The probability of helical structure is shown as positive values, while negative values are used for the probability of an elongated structure

carbohydrate binding modules. The typical β-sheet scaffold support the evolution of a variety of binding specificities and affinities in xylan-specific CBMs which merits further analyses.

Acknowledgements This work was financed by SO-funds from NTNU, Norwegian University of Science and Technology and by the Norwegian NMR Platform and the KIFEE 2016–2018 program both from the Research Council of Norway (Grant Numbers 226244 and 249797), as well as Graduate School DTU Scholarship, Lyngby, Denmark and the Danish Research Council for Independent Research, Natural Sciences (DFF, FNU) by a Research Project 2 grant (Grant ID: 4002-00297B).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The authors declare that the experiments described in this publication were done in compliance with the ethical standards of the countries in which they were performed.

References

Chassard C, Goumy V, Leclerc M, Del'homme C, Bernalier-Donadille A (2007) Characterization of the xylan-degrading microbial community from human faeces. FEMS Microbiol Ecol 61:121–131. https://doi.org/10.1111/j.1574-6941.2007.00314.x

Donohoe DR, Collins LB, Wali A, Bigler R, Sun W, Bultman SJ (2012) The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. Mol Cell 48:612–626. https://doi.org/10.1016/j.molcel.2012.08.033

Keller R (2004) The computer aided resonance assignment (Goldau, Switzerland C Verlag (Ed)

Koropatkin NM, Cameron EA, Martens EC (2012) How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol 10:323–335. https://doi.org/10.1038/nrmicro2746

Leth ML et al (2018) Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. Nat Microbiol 3:570–580. https://doi.org/10.1038/s41564-018-0132-8

Louis P, Flint HJ (2009) Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine FEMS. Microbiol Lett 294:1–8. https://doi.org/10.1111/j.1574-6968.2009.01514.x

Martens EC et al (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol 9:e1001221. https://doi.org/10.1371/journal.pbio.1001221

Mirande C, Kadlecikova E, Matulova M, Capek P, Bernalier-Donadille A, Forano E, Bera-Maillet C (2010) Dietary fibre degradation and fermentation by two xylanolytic bacteria Bacteroides xylanisolvens XB1A and *Roseburia intestinalis* XB6B4 from the human intestine. J Appl Microbiol 109:451–460. https://doi.org/10.1111/j.1365-2672.2010.04671.x

Morrison DJ, Preston T (2016) Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes 7:189–200. https://doi.org/10.1080/19490976.2015.1134082

Rennie EA, Scheller HV (2014) Xylan biosynthesis. Curr Opin Biotechnol 26:100–107. https://doi.org/10.1016/j.copbio.2013.11.013

Rogowski A et al (2015) Glycan complexity dictates microbial resource allocation in the large intestine. Nat Commun 6:7481. https://doi.org/10.1038/ncomms8481

Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR 56:227–241. https://doi.org/10.1007/s10858-013-9741-y

Sheridan PO et al (2016) Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. Microb Genomics 2:e000043. https://doi.org/10.1099/mgen.0.000043

Xu S, Liu CX, Xu W, Huang L, Zhao JY, Zhao SM (2017) Butyrate induces apoptosis by activating PDC and inhibiting complex I through SIRT3 inactivation. Signal Transduct Target Ther 2:16035. https://doi.org/10.1038/sigtrans.2016.35

Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25:173–195. https://doi.org/10.1023/a:1022836027055

Zhang M et al (2014) Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc Natl Acad Sci USA 111:E3708–E3717. https://doi.org/10.1073/pnas.1406156111

Chapter 4 Structure and functional insight into a new xylan binding module of the multi-modular xylanase from *Roseburia intestinalis* (Manuscript in preparation)

The present chapter is a manuscript in preparation, which is envisioned to include additional studies. It is of interest to identify unique properties of the CBMx (reported in chapter 2 and 3) and to assign it into a new CBM family. The manuscript includes a solved crystal structure of CBMx and its binding to xylans and decorated XOS are explored by NMR. The NMR results demonstrate structural changes of CBMx upon binding and that the preferred ligand is WAX. Further analysis of the NMR data will explore the dynamics of the protein. Additional experiments planned to be included in the manuscript are; mutations studies of crucial residues in the binding site and cryo-EM of the full length GH10. The crystallization was performed by Morten Ejby and the NMR experiments were performed by Finn Lillelund Aachmann and his group. I contributed to the presented study by writing the manuscript and by cloning of the CBMx, together with establishment of protein expression and purification conditions.

Supplementary information is included in the end of the chapter

Structure and functional insight into a new xylan binding module of the multi-modular xylanase from *Roseburia intestinalis*

Maria Louise Leth[1*], Morten Ejby[1*], Eva Madland[2], Yoshihito Kitaoku[2], Dirk Jan Slotboom[3], Albert Guskov[3], Finn Lillelund Aachmann[2], Maher Abou Hachem[1]

1. Dept. of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. 2. NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, N-7491 Trondheim, Norway. 3. Membrane Enzymology, Institute for Biomolecular Sciences & Biotechnology, Rijksuniversiteit Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. * These authors contributed equally to this work.

Abstract

Efficient capture of glycans, the prime metabolic resource in the human gut, confers a key competitive advantage for gut microbiota members equipped with extracellular glycoside hydrolases (GHs) to target these substrates. The association of glycans to the bacterial cell surface is typically mediated by carbohydrate binding modules (CBMs). Here we report the structure of *Ri*CBMx, a representative of a new family of xylan binding CBMs present in xylanases from human gut Clostridiales that correlate with a healthy microbiota. *Ri*CBMx of the GH10 xylanase from *Roseburia intestinalis* adopts a canonical β-sandwich fold, but shows structural divergence from known CBMs. The open and shallow ligand-binding site of *Ri*CBMx is more similar to CBMs that target crystalline substrates (Type-A) than counterparts that bind glycan chains in deep clefts (type-B). Curiously, *Ri*CBMx recognizes only a single xyloxyl ring with direct polar bonds, which is unprecedented amongst previously reported type-B CBMs that display more extensive and less focused polar ligand recognition patterns. The architecture of *Ri*CBMx is consistent with an atypically low binding affinity compared to most xylan binding CBMs. Titration analyses using NMR spectroscopy indicated that ligand binding induces conformational re-arrangements and that *Ri*CBMx prefers arabinoxylan over glucuronoxylan, consistent with the apolar-negativley charged surface flanking the binding site. Altogether, this study provides insight into the structural features that shape low-affinity CBMs, which are intermediate of type A and type B.

Introduction

The gut microbiota (HGM) consists of trillions of microorganisms that exert a profound impact on human health, especially via modulation of host immune- and metabolic homeostasis[1,2]. The molecular dialogue of the microbiota with the host is typically communicated via microbial metabolites, whereby short chain fatty acids (SCFAs) produced from fiber fermentation play a key role[3]. The most common SCFAs are acetate, propionate and butyrate, all of which are considered beneficial to human health[4]. Notably, the fiber fermentation SCFA profiles generated are specific to distinct taxonomic groups, for example members of the dominant genus *Bacteroides* produce mainly acetate (and lower amounts of propionate) , whereas members from *Clostridium* group XIVa group[5,6] are key butyrate producers[7]. Butyrate has received increasing attention due to its role in enforcing the gut barrier by increasing the proliferation rate of colonocytes and strengthening tight junctions. Moreover, butyrate down-regulates the expression of inflammatory cytokines and increases colonic regulatory T cells by via inhibition of host histone deacetylases[8,9]. Thus, butyrate producers are considered an indicator of a healthy HGM and seem to make a marked contribution to maintaining a balanced and healthy community in the human gut[10]. Despite these pronounced physiological roles, little attention has been dedicated to understating the interactions of butyrate producing members of the HGM with dietary glycans, as opposed to other taxonomic groups that are ascribed a probiotic status, e.g. bifidobacteria[11–13] and lactobacilli[14,15].

Roseburia intestinalis from the *Clostridium* cluster XIVa group of Firmicutes is an abundant (up to 5 % of the total microbiota) and prevalent butyrate producer[7,16]. The abundance of *R. intestinalis* is reduced in type 2 diabetes[17], Chron's disease[18–20], and colorectal cancer[21] patients, which is consistent with the association of this species to a balanced microbiota in healthy humans. *R. intestinalis* has also been shown to adhere to mucin[22], reflecting intimate association with the host and production of butyrate close to the surface of the enterocytes. *R. intestinalis* is atypical amongst other human gut Firmicutes by encoding a considerable repertoire (>130) of glycoside hydrolases and polysaccharide lyases[23] indicative of extensive saccharolytic potential. Accordingly, *R. intestinalis* is an appropriate model to investigate the strategy of complex glycan utilisation by butyrate producing *Clostridium* XIVa members.

R. intestinalis and *Eubacterium* rectale, both affiliated to the *Clostridium* XIVa cluster of Firmicutes, have been proposed to be key primary degraders of the structural glycan xylan based on enrichment from faecal samples and *in vitro* growth on xylan[24,25], which is one of the most abundant glycans in human diet. Xylan comprises a β-(1⟶4)-xyloxyl backbone with a variety of sidechain substitutions that

vary considerably according to botanic origin and tissue. Arabinoxylan (AX), the dominant structural component in the cereal cell wall[26], is substituted with L-arabinosyl residues at C2, C3 or both positions of backbone xyloxyl units. Xylan is also present in lower amounts in vegetables and fruits as glucuronoxylan[27] (GX), which is decorated with (4-*O*-methyl)glucuronic acid at the C2 position of xylosyl units. Both AX and GX are further acetylated at C2, C3 or both positions. The molecular apparatus of xylan utilisation by *R. intestinalis* has been recently described[5]. Extracellular capture and break down of xylan is mediated by a modular xylanase of GH10 (*Ri*GH10A). This enzyme, which is conserved within the species, comprises an N-terminal carbohydrate binding module (CBM) from a previously unknown family (henceforth designated as *Ri*CBMx) followed by a CBM22, a GH10 catalytic module, a tandem repeat of CBM9 and two C-terminal putative cell-attachment domains. Curiously, *Ri*CBMx was specific to xylan, but it displayed relatively low affinity (K_D≈0.5 mM for xylohexaose (X6) compared to about 7-fold higher average affinity of the truncated enzyme lacking this CBM for the same ligand. Interestingly, *Ri*CBMx prefers the nutritionally more abundant arabinoxylan as compared to glucuronoxylan judged by retardation in affinity electrophoresis gels[5].

Association to complex glycans, such as xylan, offers a competitive advantage for bacteria in the competitive milieu of the gut. Firmicutes from *Clostridium* XIVa cluster frequently have large modular cell-attached glycoside hydrolase (GHs) containing multiple carbohydrate binding modules (CBMs) for capture and hydrolysis of polysaccharides[5,6,28,29]. To examine the mode of recognition and discrimination of *Ri*CBMx to different xylans as well as the contribution of this domain to the dynamic binding of the appended xylanase *Ri*GH10A, we have determined the structure of this module and performed binding analyses to glucurono- and arabinoxylan and oligosaccharides thereof using NMR spectroscopy.

*Ri*CBMx shares a β-sandwich fold with other CBMs, but displays very low structural similarity to functionally characterized orthologues, which merit its assignment in a new CBM that currently represents about 19 sequences from taxonomically related human gut Lachnospiraceae. The ligand-binding site of *Ri*CBMx is open and shallow, with only three residues recognising a single xylosyl unit out of four observed in the crystal structure in complex with X6. The binding site architecture and the focused direct polar interactions to a single saccharide ring are atypical for type B CBMs that frequently possess deep clefts and more extensive polar ligand contacts along their binding sites. These features are discussed in relation to the measured affinity and in comparison to other described xylan-binding modules, especially those occurring in GH10 endo-xylanases.

Results

Crystal structure

We determined the structure of *Ri*CBMx in complex with xylohexaose (X6). The structure was solved in the hexagonal space group $P6_5$ (with 6 molecules in the asymmetric unit) using single-wavelength anomalous diffraction (SAD) with the experimental phase information obtained from data collected on crystals soaked with Tb-Xo4[30] using the Tb anomalous scattering for phasing. The data collection and refinement statistics are shown in Table 1. The structure of *Ri*CBMx was solved to 1.9 Å resolution revealing a β-sandwich fold, consisting of two sheets formed by 11 antiparallel β-strands and 2 helical turns (right handed 3_{10}-helices) connected by loops (Fig. 1A). β-Sheet 1 forms the concave face of the β-sandwich and consists of the strands β2(K39-G43), β5(Y62-T68), β7(I92-Y97), β8(T108-L112) and β10(D129-I135). β-Sheet 2 is formed by β1(V29-T34), β3(D46-A50), β4(G53-F58), β6(N79-A86), β9(E117-I120) and β11(A143-L154). A striking feature of the CBM is the open solvent accessible ligand-binding site (described in more detail below) that runs almost orthogonal to the β-strands of sheet 1 (Fig. 1A). A search of the DALI server against the protein data bank (PDB) suggested that the closest structural relative of CBMx is the CBM29.2, from the anaerobic fungus *Piromyces equi*[31] (1W9F, *Z*-score=12.8, primary structure identity 12%), which has specificity for β-mannan[32]. The second closest structural hit is a CBM84 from a from a xanthan lyase family 8 from *Paenibacillus nanensis*[33]. Although the overall structural fold is shared between these modules, the low (<12%) shared sequence identity and the divergence of the binding sites (none of the residues that provide aromatic stacking interactions to the ligand are conserved) justify the assignment of *Ri*CBMx as a representative of a new CBM family.

Ligand binding site

The crystal structure of *Ri*CBMx in complex with X6 shows relatively densities for four xylosyl units. The ligand-binding site features an open and shallow surface with the ligand bound in a relaxed helical conformation[34](Supplementary 1). The ligand-binding site is defined by Y62, which stacks onto the terminal reducing end moiety of the xylosyl, which defines position 1 (Fig. 1B). Xylo-oligosaccharide (XOS) ligands can, however, be accommodated in the opposite directionality with equivalent contacts (non-reducing end xylosyl stacking onto Y62, Supplementary 2). Our description will focus on the former orientation for clarity. The second aromatic ridge is provided by Y110 that stacks onto the xylosyl unit at position 3. A third potential stacking residue is W42, which is resembles one side of a sugar tong (Fig. 1C), but the indole face is blocked by a sidechain from a neighboring molecule in the crystals. The recognition of the helical conformation of the XOS, is facilitated by the planes of the

aromatic rings of Y62 and Y110 being almost orthogonal (≈100˚) to each other (Fig. 1B). The only direct potential polar interactions are observed at position 3 (Xyl3) between the C2-OH and K95 N^{ζ}, C3-OH and -Q64 $N^{\varepsilon 2}$, and K95 N^{ζ} and D102 $O^{\delta 2}$ (Supplementary 2). Additional water mediated potential hydrogen bonds may also contribute to the recognition (Supplementary 2).

Table 1. Data collection and refinement statistics

	CBMx
Beamline	PETRA III P13
PDB ID	-
Wavelength (Å)	1.649
Resolution range (Å)	70.9 - 1.9 (1.98 - 1.91)
Space group	$P6_5$
Unit cell	141.87 141.87 60.6 90 90 120
Unique reflections [a]	53405 (5006)
Multiplicity [a]	9.6 (6.8)
Completeness (%)[a]	99.30 (93.65)
CC½[a]	0.997 (0.898)
Mean I/σ(I) [a]	14.48 (3.21)
Wilson B-factor	19.18
R-factor [b]	0.2059 (0.2095)
R-free [b,c]	0.2522 (0.2640)
Number of atoms	6885
Macromolecules	5954
Ligands	282
Water	649
Protein residues	786
RMS bonds (Å)	0.011
RMS angles (°)	1.43
Ramachandran favored (%)	98.19
Ramachandran outliers (%)	0
Clash score	3.60
Average B-factor	20.54
Macromolecules	19.58
Ligands	29.99
Water	25.25

[a] Values in the parenthesis are for the highest resolution shell.
[b] Values in the parenthesis are for before refinement.
[c] Refinement is incomplete as the model will be moved to a higher resolution dataset and the Tb-Xo4 and other solvent has not been finalized in this model.

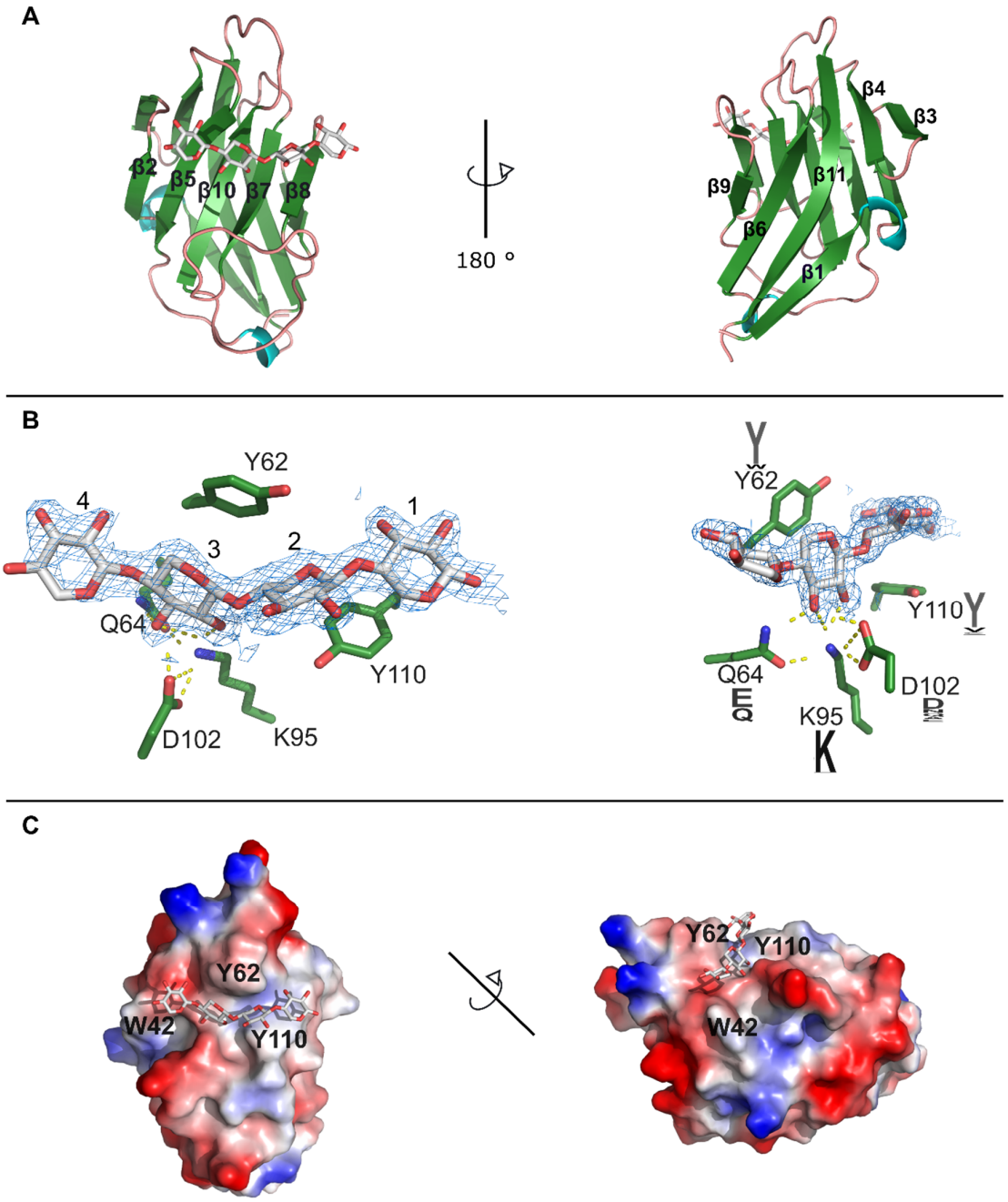


Figure 1. Crystal structure of *Ri*CBMx. A) Cartoon model of β-sandwich structure of *Ri*CBMx. The left panel is a top view of sheet 1 formed by five β-strands. The four visible rings of the soaked xylohexoase ligand are shown in sticks. The view is rotated 180˚ in the right panel to show sheet 2 formed by six β-strands. B) The left panel is a close-up of the ligand bindings site with subsites numbered in Arabic numerals starting from the reducing end at position 1. The two aromatic residues Y110 and Y62 that stack onto xylosyl rings at positions 1 and 3, respectively. The only direct polar interactions that recognize the C2 and C3 hydroxyl groups the xylosyl group at position 3 are shown and the $2F_{obs}–DF_{calc}$ electron map for the bound ligand is shown at a contour level of 1σ (blue mesh). The right panel is the binding site in position 1 from another angle and the conservation of the binding residues is shown for clarity. C) The electrostatic potential of *Ri*CBMx (at pH=7) is shown to highlight the topology and the chemistry of the ligand binding site of *Ri*CBMx. The two aromatic stacking residues Y62 and Y110 are shown in add on to the potential aromatic binding residue W42.

Ligand binding analysis using NMR spectroscopy

The changes in ^{15}N-HSQC (Heteronuclear single quantum coherence spectroscopy) spectra of *Ri*CBMx were monitored and the change in chemical shifts for both the N and H atoms upon titration with undecorated xylotetraose (X4), a 1:1 mixture of 3^3-α-L- and 2^3-α-L-arabinofuranosyl-xylotetraose (XAXXX) and 2^3-(4-*O*-methyl-α-D-glucuronyl)-xylotetraose (XUXXX) was followed. Here we describe the preliminary analysis of the result. The affinity of the *Ri*CBMx was lowest for XUXXX, while the higher affinity for XAXXX and X4 resulted in a chemical shift difference in the same order (Fig. 2, Table 2). The change in chemical shift occurred mainly at the binding site and the flanking area (Fig. 2). The amino acids Y62, Q64, K95, D102 and Y110, which are observed to interact with the ligand in the crystal structure, showed a significant chemical shift difference after titration with the three ligands, except for Q64 with XUXXX. An interesting observation is that G111 undergoes a change in chemical shift in the ^{1}H dimension only for the decorated substrates, which is suggestive this region of CBMx being involved in the accommodation of ligand units not observed in the crystal structure. Neighboring G111, is Y110 which provides aromatic stacking interactions for substrate.

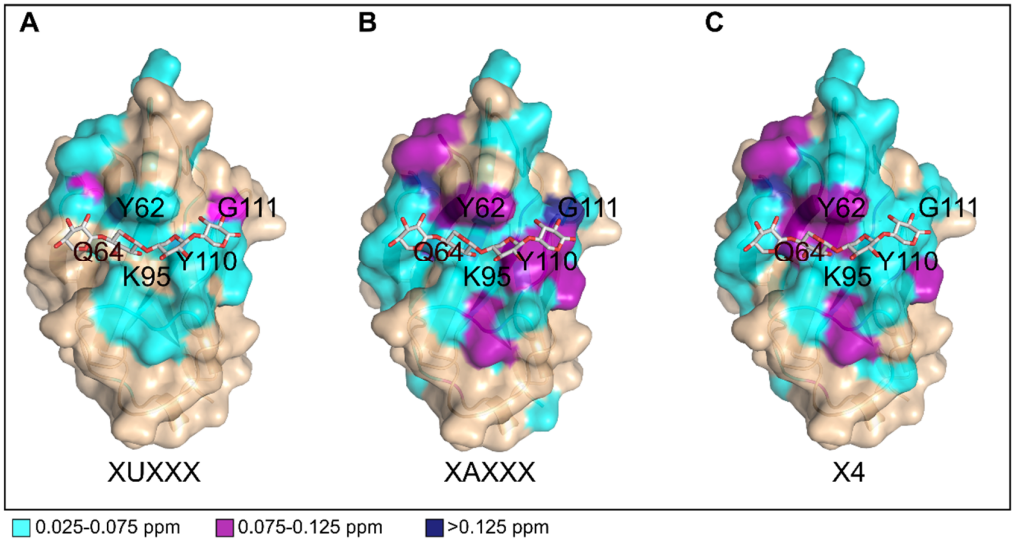


Figure 2 Chemical shift differences for *Ri*CBMx. The chemical shift differences are after titration with xylo-oligosaccharides; glucurono-xylotetraose (XUXXX), α-L-arabinofuranosyl-xylotetraose (XAXXX) and xylotetraose (X4).

Table 2 Binding parameters determined by NMR

	K_D (mM)	B_{max} (Δδ at saturation)
X4	1.09	0.19
XAXXX	1.23	0.17
XUXXX	22.89	0.15

Binding parameters are calculated from a single titration experiment

The interactions between *Ri*CBMx and the polymeric substrates birch glucuronoxylan (BGX) and wheat arabinoxylan (WAX) were followed in ^{15}N-HSQC spectra upon titration (Fig. 3). Due to the strong interaction between *Ri*CBMx and WAX, some of the signals were broadened beyond detection. The signals for WAX expanded as the only tested substrate to the backside of the protein, which likely stems from an induced fit conformational change upon ligand binding. The chemical shift difference was lower for BGX, indicating weaker binding affinity to *Ri*CBMx than WAX. This, in addition to the observations made with oligomeric substrates, is evidence for that *Ri*CBMx prefers arabinosyl substitutions compared to glucurono acids substitutions both on XOS and polymeric xylan.

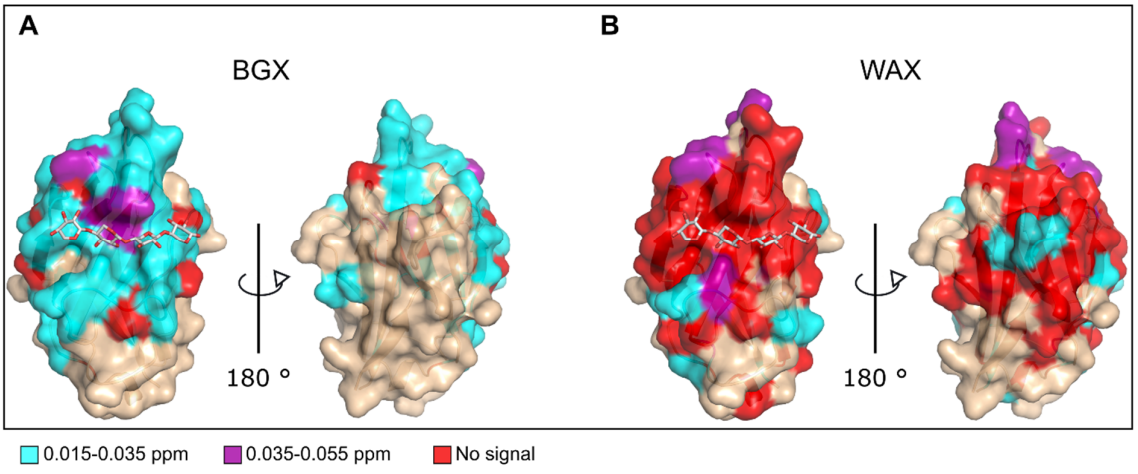


Figure 3 Chemical shift differences for *Ri*CBMx. The chemical shift differences are after titration with birch glucuronoxylan (BGX) and wheat arabinoxylan (WAX).

Close homologue of RiCBM are observed in a taxonomically related Lachnospiraceae

We have previously described that *Ri*CBMx confers affinity to xylan and xylo-oligosaccharides but lacks homologues with an assigned function[5]. A blast search against the non-redundant database identified 19 homologs from different butyrate producing strains from the Lachnospiraceae family of gut Firmicutes. An analysis of these sequences revealed that several structural residues and binding

residues are conserved (Fig. 1B, Supplementary 3,4,). This and the narrow distribution among related gut bacteria points to a highly specialized nature of these binding modules.

Discussion

In this study, we have determined the crystal structure for *Ri*CBMx in complex with X6. Despite the typical β-sandwich fold, *Ri*CBMx did not display high structural similarity to any CBM families or other characterized proteins. Notably, the closest structural homologues were CBMs with affinity to polysaccharides with a different structural symmetry than xylan, such as β-mannan or xanthan. The lack of conserved ligand binding residues between *Ri*CBMx and distant functionally described orthologues, is consistent with the functional divergence of the new CBM family represented by *Ri*CBMx. Currently, 19 non-redundant sequences are retrieved from the NCBI database. Both the aromatic and the polar residues that interact with the bound ligand in *Ri*CBMx are highly conserved in these sequences (Fig. 1B, Supplementary 3,4,).

*The architecture of the ligand-binding site of Ri*CBMx *is consistent with low affinity for xylo-oligosaccharide ligands.*

The ligand-binding site of *Ri*CBMx features an extended shallow and open binding surface that accommodates four xyloxyl units. The experimental data from our previous work show about a 4-fold increase in affinity for X6 as compared with X4, consistent with the presence of additional contacts that stabilise the binding beyond the observed X4 ligand as observed in other xylan binding modules of CBM6[35] and CBM15[36]. The solvent exposed W42 constitutes a potential stacking residue, but the space in front of the indole ring is blocked by a sidechain from a neighboring molecule in the crystal packing, which hampers evaluating the relevance of this residue in ligand binding. The chemical shift changes of this residue, however, are just above significance threshold for XAXXX and X4 and below that for the lower affinity XUXXX. Notably, W42 is conserved in all but two homologues of *Ri*CBMx, which together with the location and solvent accessibility of this residue merit further mutational analysis to evaluate possible role in ligand binding.

The architecture of the binding site of *Ri*CBMx is different from the type-B xylan specific CBMs[37] e.g. from CBM4[38], CBM6[39], CBM15[36], CBM22[40]. The deeper and more occluded binding site in these latter CBM families is defined by loops connecting the sandwich β-strands and pointing into the binding site. By contrast, the equivalent loops in *Ri*CBMx are pointing downwards and away from the ligand, which creates a relatively flat open binding surface topology (Fig. 1C) that better resembles type-A CBMs, specific to crystalline flat substrates, e.g. cellulose and chitin[37]. To our knowledge, the only characterized xylan specific CBM, which has a similar open binding site to *Ri*CBMx, is *Cf*CBM2b from

Cellulmonas fimi assigned into CBM2[41]. This latter family harbors mostly cellulose binding type A CBMs (CBM2a) and xylan specific CBM2b modules. Interestingly, the affinity of *Cf*CBM2b to X6 (K_D=0.2 mM) similarly to *Ri*CBMx is substantially lower than typical type B xylan-specific CBMs, which have affinities about 10-fold higher to X6[39,40,42]. A unique feature of *Ri*CBMx, appears to be that the direct polar interactions are recognizing only a single xylosyl unit (Supplementary 2), as opposed to counterparts from other families, where direct hydrogen bonds distributed to recognize 2−3 xylosyl rings along the binding site. Finally, the type of sidechains decorations appears to impact substrate affinity, as glucuronic acid substitution seem to reduce affinity, possibly due to electrostatic repulsion. This is consistent with the apolar or negative surface electrostatic potential around the binding site (Fig. 1C). In contrast, arabinosyl decorations are either tolerated or recognized, based on the similar affinities for the undecorated and decorated ligand X4 and the markedly higher affinity for WAX as compared to BGX (Fig. 3).

Rationale for having lower affinity xylan binding in modular xylanase?

Having large extracellular enzymes with a variety of CBMs seems to be common in Lachnospiraceae form the human gut. *R. intestinalis* has a large modular GH26 mannanase with two CBMs[6] and both *Eubacterium rectale* and *Butyrivibrio fibrisolvens* possess large modular α-amylases with 5 and 2 CBMs, respectively for capturing starch[28,43]. *Ri*CBMx is joint to a CBM22, which precedes the catalytic GH10 module that is flanked by a tandem CBM9 repeat. Notably the architecture of characterized CBM22 and CBM9 are different from each other and from the *Ri*CBMx. Members of CBM9 are type-C CBMs that possess a binding slot able to accommodate two terminal saccharide units of polysaccharides[44], whereas CBM22 possess a deep extended binding cleft[45,46]. Thus, the three different families of CBMs in *Ri*Xyn10A orchestrate the binding of substrate by being able to capture either the terminal reducing ends or internal regions of xylan by the CBM9 (assuming similar binding mode to known members) or CBM22/CBMx, respectively. These CBMs also appear to have variable affinities as judged from affinity measurements of *Ri*CBMx, the full-length enzyme and a truncated variant lacking *Ri*CBMx, which have affinities of 479 μM, 128 μM and 65 μM, respectively[5]. This variable affinity and multiplicity of CBMs may confer a dynamic binding where the substrate is anchored to the enzyme surface in between consecutive catalytic cycles to minimize diffusional loss. The evolution of low affinity CBMs may be an adaptation to increase the area of substrate binding with minimal reduction of turnover, i.e. maximizing k_{cat}/k_{off}. Additional experiments are required to evaluate the dynamics of substrate binding and translocation to *Ri*Xyn10A.

Materials and Methods

Chemicals

All chemical were of analytical grade. Wheat arabinoxylan (WAX), Xylohexaose (X6), xylotetraose (X4), 3^3-α-L- and-2^3-α-L-arabinofuranosyl-xylotetraose (XAXXX) in mixture of ~1:1 were from Megazyme (Wicklow, Ireland). 2^3-(4-*O*-methyl-α-D-glucuronyl)-xylotetraose (XUXXX) was from Cambridge Glycoscience (Cambridge, United Kingdom). Birchwood glucuronoxylan (BGX) was from Carl Roth (Karlsruhe, Germany).

Cloning

The gene fragment encoding the *Ri*CBMx from *Roseburia intestinalis* L1-82 was amplified from a plasmid encoding the full length xylanase *Ri*Xyn10A (EEVO1588.1, ROSINTL182_06494)[5] using the sense primer: 5' TTTCAGGGCGCCATGGGGGTAAAAAAAGTTTTTACTGCAGAT 3' and antisense primer: 5' GACGGAGCTCGAATTTTAATCCCCCAATTTTGCA 3'. The amplicon was cloned into the EcoRI and NcoI restriction site of a pETM-11 vector (kind gift from Dr. Gunter Stier, EMBL, Center for Biochemistry, Heidelberg, Germany)[47] using In-Fusion cloning (Takara). The construct was transformed into *Escherichia coli* DH5α and verified by full sequencing.

Expression and purification

Recombinant plasmids were transformed into BL21(DE3) (Novagen) for expression of unlabeled and $^{13}C/^{15}N$ double labeled protein and B834(DE3)(Novagen) expression selenomethionine labelled protein. Protein production was performed as previously described for unlabeled protein[5], selenomethionine labelled protein[11], and double labelled $^{13}C/^{15}N$ labelled protein used for the NMR studies[48]. Cell pellets were resuspended in buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 10% glycerol) and disrupted at 1000 bar by a single passage in a high pressure homogenizer (Standsted Fluid Power, Essex, UK). Recombinant proteins were purified from the supernatant by affinity chromatography using a 5 mL His-Trap HP column (GE Healthcare) and a standard protocol. Pure fractions were concentrated and loaded onto a Hiload 16/60 Superdex 75 pg size exclusion chromatography column (GE Healthcare) mounted on an ÄKTA-AVANT chromatograph (GE Healthcare). For crystallization the His-tag was removed using a TEV-protease. This was done by buffer exchange into buffer (50 mM Tris-HCL pH 8.0, 0.5 mM EDTA, 1 mM DTT) and next adding TEV-protease in a ratio of 1:100 (v/v). After incubation for 24 hours at room temperature, the mixture was passes through a His-Trap column, and the flow through containing the cleaved protein dialyzed into buffer (20 mM MES pH 6.5, 150 mM NaCl). Protein purity was determined by SDS-PAGE and protein concentration were measured

spectrophotometrically and calculated from the theoretical molar extinction coefficient (ε^{280nm}= 26930 and 23950 M^{-1} cm^{-1}, for tagged and non-tagged proteins, respectively).

Crystallization and structure determination

Crystals were only obtained in the presence of 1 mM X6 by vapor diffusion in hanging or sitting drops, and grew for 2 days at 5 °C with a 1:1 ratio of the protein (18 mg mL^{-1} in 10 mM MES pH 6.5 and 150 mM NaCl) and reservoir solution (0.2 M Cadmium chloride hemi(pentahydrate) 0.1 M Sodium acetate pH 4.8 and PEG 400 35 % v/v). An initial crystallisation condition (0.1 M Cadmium chloride hemi(pentahydrate), 0.1 M Sodium acetate pH 4.6 and PEG400 30 % v/v at 5 °C) was identified with the Structure Screen (Molecular Dimensions Ltd, UK), using a Mosquito® liquid handling robot (TTP Labtech, UK). The crystals were flash frozen in liquid nitrogen without cryo-protectant. Diffraction data were collected to a maximum resolution of 1.9 Å, at the DESY beamlines, Hamburg, Germany. The dataset was processed with XDS[49]. The structure was solved in the hexagonal space group $P6_5$ using single-wavelength anomalous diffraction (SAD) with the experimental phase information obtained from data collected at 7.575 KeV for crystals soaked for 1 min with 100 mM Tb-Xo4[30] (Molecular Dimensions) using the Tb anomalous scatterer for phasing. Experimental phasing, initial model building and refinements were performed in the Phenix software suite[50]. Manual structure improvement was done in Coot[51]. Ligand molecules were included after the protein parts were build and water molecules were added with Coot or manually. The overall quality of all models was checked using MolProbity[52]. The data collection and refinement statistics are presented in Table 1. The PyMOL Molecular Graphics System, Version 2.0.6 Schrödinger, LLC was used to explore the models and for rendering.

NMR spectroscopy

NMR spectra of ≈0.10 mM *Ri*CBMx in 50 mM sodium phosphate buffer pH 6.5 and 10 % D_2O were recorded at 25 °C on a Bruker Ascend 800 MHz spectrometer Avance III HD (Bruker Biospin) equipped with a 5 mm Z-gradient CP-TCI (H/C/N) Cryoprobe at the NV-NMR-Centre/Norwegian NMR Platform at NTNU.

A single NMR titration was preformed with three oligomeric substrates: X4, XAXXX or XUXXX. Titration points for X4: 0.5 mM, 1.0 mM, 2.5 mM, 5 mM and 10 mM; XAXXX: 0.2 mM, 0.5 mM, 1.0 mM, 2.5 mM, 5.0 mM and 10 mM; XUXXX: same as for XAXXX with the addition of the following four points of 12.5 mM, 15.0 mM, 20.0 mM and 25.0 mM. In addition, NMR titrations were also carried out with two xylans: BGX and WAX. The titration with BGX was performed with nine concentrations within 0.04−1.0 mg mL^{-1} and a final point at 2.0 mg BGX. For WAX eight concentrations within 0.04−0.73 mg·mL^{-1} and

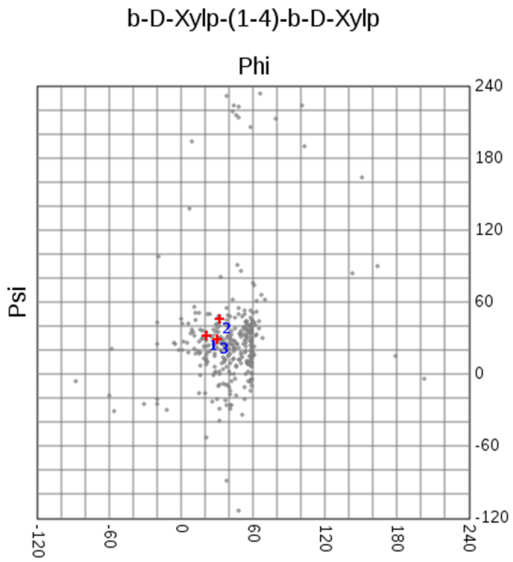
a final point of 1.4 mg WAX. 1D and ^{15}N-HSQC spectra were recorded for each titration point and processed with Topspin version 3.5 and CARA version 1.5 using backbone and side-chain assignments of CBMx have been published elsewhere[48]. The chemical shift perturbation upon titration was followed in ^{15}N-HSQC. Binding parameters were estimated by Gnuplot 5.2 (www.gnuplot.info) using an average of the chemical shift difference ($\Delta\delta$) from the titration of three amino acids, K_D X4 (A59, N63, N93), K_D XAXXX (N63, N93, G111) and K_D XUXXX (N63, N93, G111).

References

1. Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* 1–10 (2015).
2. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism. *Nature* **535,** 56–64 (2016).
3. Ríos-Covián, D. *et al.* Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* **7,** 1–9 (2016).
4. Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P. & Macfarlane, G. T. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28,** 1221–1227 (1987).
5. Leth, M. L. *et al.* Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* **3,** 570–580 (2018).
6. Rosa, S. *et al.* Molecular and Biochemical Basis for Complex β-Mannans Catabolism by the Human Gut Firmicutes Bacterium *Roseburia intestinalis*. *Nat. Commun.* (2019).
7. Louis, P. & Flint, H. J. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol. Lett.* **294,** 1–8 (2009).
8. Chang, P. V, Hao, L., Offermanns, S. & Medzhitov, R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl. Acad. Sci.* **111,** 2247 LP-2252 (2014).
9. Smith, P. M. *et al.* The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science (80-.).* **341,** 569 LP-573 (2013).
10. Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12,** 661–672 (2014).
11. Ejby, M. *et al.* An atp binding cassette transporter mediates the uptake of α-(1,6)-linked dietary oligosaccharides in bifidobacterium and correlates with competitive growth on these substrates. *J. Biol. Chem.* **291,** 20220–20231 (2016).
12. Viborg, A. H. *et al.* A β1-6/β1-3 galactosidase from *Bifidobacterium animalis* subsp. lactis Bl-04 gives insight into sub-specificities of β-galactoside catabolism within Bifidobacterium. *Mol. Microbiol.* **94,** 1024–1040 (2014).
13. Morrill, J. *et al.* The GH5 1,4-β-mannanase from *Bifidobacterium animalis* subsp. lactis Bl-04 possesses a low-affinity mannan-binding module and highlights the diversity of mannanolytic enzymes. *BMC Biochem.* **16,** 26 (2015).
14. Møller, M. S. *et al.* Enzymology and Structure of the GH13_31 Glucan 1,6-α-Glucosidase That Confers Isomaltooligosaccharide Utilization in the Probiotic *Lactobacillus acidophilus* NCFM. *J. Bacteriol.* **194,** 4249 LP-4259 (2012).
15. Møller, M. S. *et al.* An Extracellular Cell-Attached Pullulanase Confers Branched α-Glucan Utilization in Human Gut *Lactobacillus acidophilus*. *Appl. Environ. Microbiol.* **83,** e00402-17 (2017).
16. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
17. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).
18. Erickson, A. R. *et al.* Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. *PLoS One* **7,** (2012).
19. Willing, B. P. *et al.* A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139,** 1844–1854.e1 (2010).
20. Shen, Z. *et al.* Insights into *Roseburia intestinalis* which alleviates experimental colitis pathology by inducing anti-inflammatory responses. *J. Gastroenterol. Hepatol.* **33,** 1751–1760 (2018).
21. Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *ISME J.* **6,** 320–329 (2012).
22. Van den Abbeele, P. *et al.* Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an *in vitro* gut model. *ISME J.* **7,** 949–61 (2013).
23. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11,** 497–504 (2013).
24. Chassard, C., Goumy, V., Leclerc, M., Del'homme, C. & Bernalier-Donadille, A. Characterization of the xylan-degrading microbial community from human faeces. *FEMS Microbiol. Ecol.* **61,** 121–131 (2007).
25. Sheridan, P. O. *et al.* Polysaccharide utilisation loci and nutritional specialisation in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genomics* **2,** (2016).
26. Izydorczyk, M. S. & Biliaderis, C. G. Cereal arabinoxylans: advances in structure and physicochemical properties. *Carbohydr. Polym.* **28,** 33–48 (1995).
27. Selvendran, R. R. Chemistry of plant cell walls and dietary fibre. *Scand. J. Gastroenterol.* **5521,** 33–41 (1987).
28. Ramsay, A. G., Scott, K. P., Martin, J. C., Rincon, M. T. & Flint, H. J. Cell-associated α-amylases of butyrate-producing Firmicute bacteria from the human colon. *Microbiology* **152,** 3281–3290 (2006).
29. Cockburn, D. W. *et al.* Molecular details of a starch utilization pathway in the human gut symbiont Eubacterium rectale. *Mol. Microbiol.* **95,** 209–230 (2015).
30. Engilberge, S. *et al.* Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chem. Sci.* **8,** 5909–5917 (2017).
31. Flint, J. *et al.* Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules. *J. Biol. Chem.* **280,** 23718–26 (2005).
32. Freelove, A. C., Bolam, D. N., White, P., Hazlewood, G. P. & Gilbert, H. J. A novel carbohydrate-binding protein is a

component of the plant cell wall-degrading complex of *Piromyces equi*. *J. Biol. Chem.* **276,** 43010–7 (2001).
33. Jensen, P. F. *et al.* Structure and Dynamics of a Promiscuous Xanthan Lyase from Paenibacillus nanensis and the Design of Variants with Increased Stability and Activity. *Cell Chem. Biol.* (2018). doi:10.1016/J.CHEMBIOL.2018.10.016
34. Lütteke, T., Frank, M. & von der Lieth, C.-W. Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res.* **33,** D242–D246 (2005).
35. Pires, V. M. R. *et al.* The crystal structure of the family 6 carbohydrate binding module from Cellvibrio mixtus endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *J. Biol. Chem.* **279,** 21560–8 (2004).
36. Szabó, L. *et al.* Structure of a family 15 carbohydrate-binding module in complex with xylopentaose: Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* **276,** 49061–49065 (2001).
37. Gilbert, H. J., Knox, J. P. & Boraston, A. B. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol* **23,** (2013).
38. Simpson, P. J. *et al.* The Solution Structure of the CBM4-2 Carbohydrate Binding Module from a Thermostable *Rhodothermus marinus* Xylanase. *Biochemistry* **41,** 5712–5719 (2002).
39. Czjzek, M. *et al.* The Location of the Ligand-binding Site of Carbohydrate-binding Modules That Have Evolved from a Common Sequence Is Not Conserved. *J. Biol. Chem.* **276,** 48580–48587 (2001).
40. Xie, H. *et al.* Clostridium thermocellum Xyn10B carbohydrate-binding module 22-2: The role of conserved amino acids in ligand binding. *Biochemistry* **40,** 9167–9176 (2001).
41. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J. & Williamson, M. P. The Structural Basis for the Ligand Specificity of Family 2 Carbohydrate-binding Modules. *J. Biol. Chem.* **275,** 41137–41142 (2000).
42. Sainz-Polo, M. A., González, B., Menéndez, M., Pastor, F. I. J. & Sanz-Aparicio, J. Exploring multimodularity in plant cell wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem. *J. Biol. Chem.* **290,** 17116–17130 (2015).
43. Cockburn, D. W. *et al.* Novel carbohydrate binding modules in the surface anchored α-amylase of Eubacterium rectale provide a molecular rationale for the range of starches used by this organism in the human gut. *Mol. Microbiol.* **107,** 249–264 (2018).
44. Notenboom, V., Boraston, A. B., Kilburn, D. G. & Rose, D. R. Crystal structures of the family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10a in native and ligand-bound forms. *Biochemistry* **40,** 6248–6256 (2001).
45. Najmudin, S. *et al.* Putting an N-terminal end to the Clostridium thermocellum xylanase Xyn10B story: Crystal structure of the CBM22-1-GH10 modules complexed with xylohexaose. *J. Struct. Biol.* **172,** 353–362 (2010).
46. Charnock, S. J. *et al.* The X6 'thermostabilizing' domains of xylanases are carbohydrate-binding modules: Structure and biochemistry of the Clostridium thermocellum X6b domain. *Biochemistry* **39,** 5013–5021 (2000).
47. Dümmler, A., Lawrence, A.-M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.* **4,** 34 (2005).
48. Madland, E. *et al.* 1H, 13C and 15N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from *Roseburia intestinalis*. *Biomol. NMR Assign.* (2018). doi:10.1007/s12104-018-9850-3
49. Kabsch, W. *XDS*. *Acta Crystallogr. Sect. D* **66,** 125–132 (2010).
50. Adams, P. D. *et al. PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66,** 213–221 (2010).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. Sect. D* **66,** 486–501 (2010).
52. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66,** (2010).
53. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
54. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–W324 (2014).

Supplementary 1



1:

b-D-Xylp-(1-4)-b-D-Xylp-(1-4)-**b-D-Xylp**-(1-4)-**b-D-Xylp**

Linkage Path: [4]
PDB Residues: XYP1I,XYP2I
phi: 21.4, psi: 32.3

2:

b-D-Xylp-(1-4)-**b-D-Xylp**-(1-4)-**b-D-Xylp**-(1-4)-b-D-Xylp

Linkage Path: [4 4]
PDB Residues: XYP2I,XYP3I
phi: 32.7, psi: 46.6

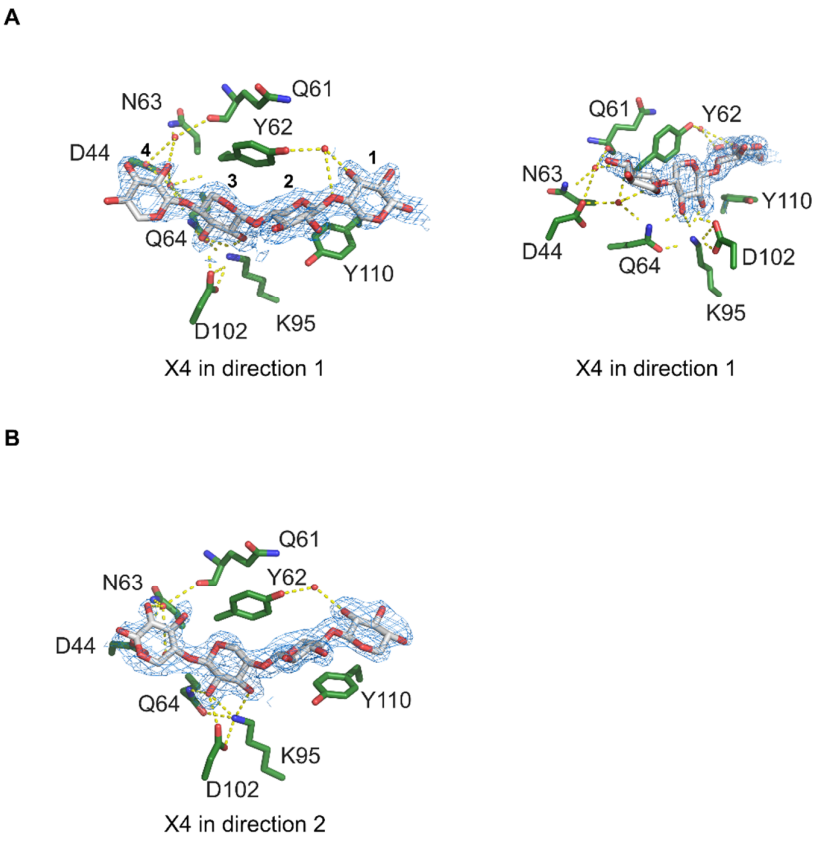
3:

b-D-Xylp-(1-4)-**b-D-Xylp**-(1-4)-b-D-Xylp-(1-4)-b-D-Xylp

Linkage Path: [4 4 4]
PDB Residues: XYP3I,XYP4I
phi: 30.2, psi: 29.6

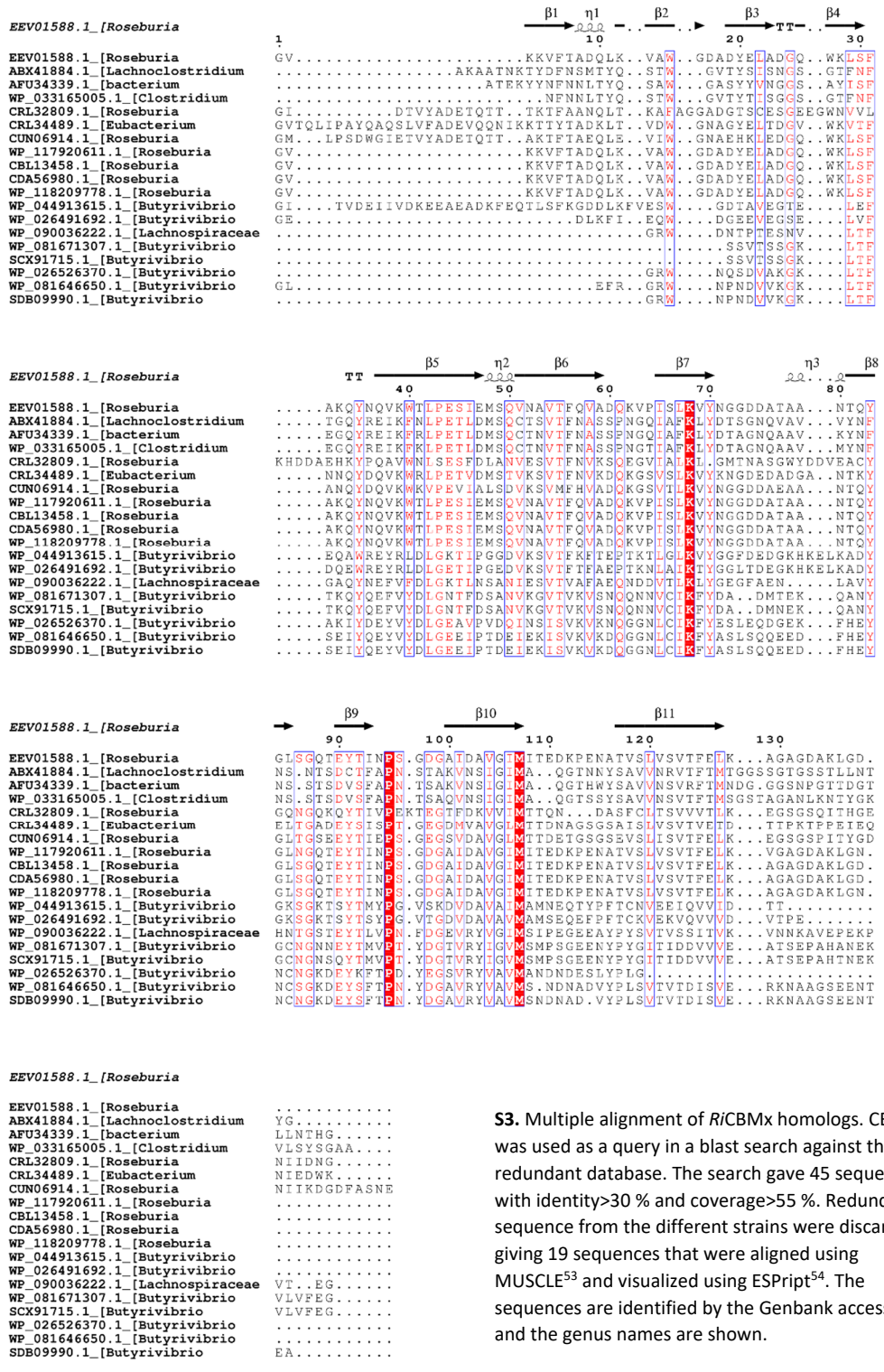
S4: CArbohydrate Ramachandran Plot (CARP) output[34]. The output shows that the ligand has a relaxed helical conformation

Supplementary 2



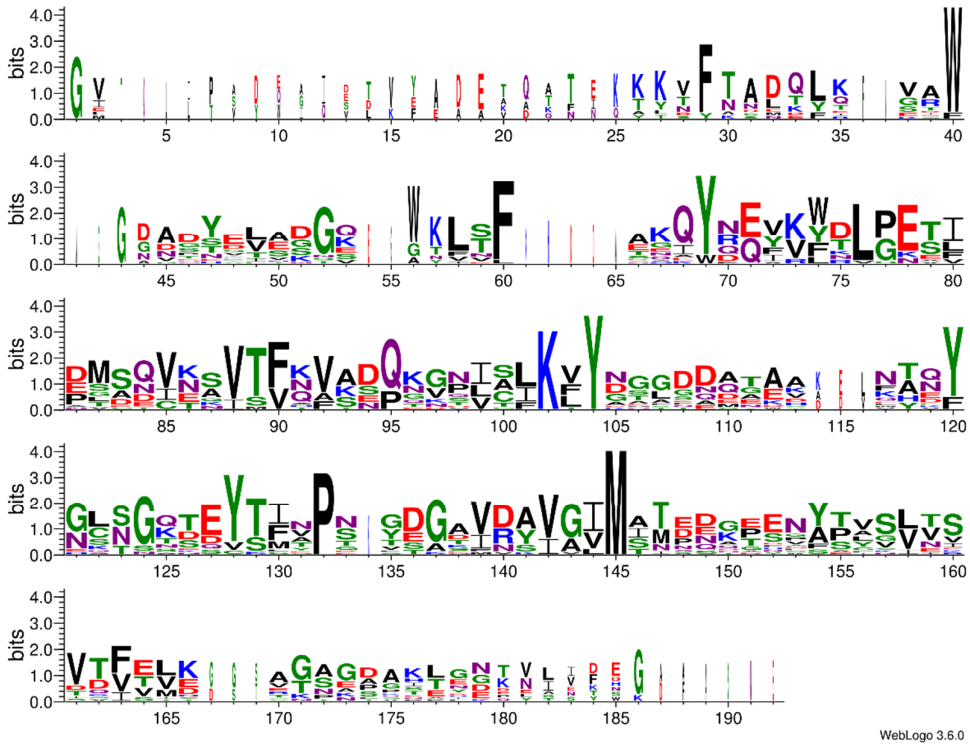
S2. Close-up of the binding site. A) Xylotetraose bound in directionality 1 with the reducing end bound at position 1. Both the direct and water mediated potential hydrogen bonds are shown in addition to the two aromatic stacking residues. B) Same as A, but with the xylotetraose bound with the opposite directionality.

Supplementary 3



S3. Multiple alignment of *Ri*CBMx homologs. CBMx was used as a query in a blast search against the non-redundant database. The search gave 45 sequences with identity>30 % and coverage>55 %. Redundant sequence from the different strains were discarded, giving 19 sequences that were aligned using MUSCLE[53] and visualized using ESPript[54]. The sequences are identified by the Genbank accession and the genus names are shown.

Supplementary 4



S4: Consensus logo for CBMx. The sequence logo was created with the multiple alignment of *Ri*CBMx and visualized using WebLogo 3.6 (weblogo.berkeley.edu).

Chapter 5 Paper 3: Molecular and Biochemical Basis for Complex β-Mannans Catabolism by the Human Gut Firmicutes Bacterium *Roseburia intestinalis*

The present paper provides insight into the molecular machinery used by the butyrate producing human gut symbiont *R. intestinalis* to utilize dietary mannan. A transcriptomic study on mannans was performed in parallel with the study on xylan in chapter 2. The transcriptomic data revealed two gene loci responsible for mannan utilization and is the initial foundation for the study. In the paper a proteomics study confirms the transcriptomic results and detailed characterization of mannan degradation and uptake apparatus are presented. In this paper *R. intestinalis* demonstrated that it is able to compete for mannan with *B. ovatus in vitro* and curiously, *in vivo* experiment showed that β-mannan can increase the number of *R. intestinalis* in mice after one day, together with other fiber degrading species, whereas the counts of mucolytic taxonomic group are significantly reduced. The data and analysis for the presented paper are mainly performed by Sabina Leanti La Rosa together with Bjørge Westereng and members of his group. Mice experiments were performed by Eric Martens and his group. Initial establishment of growth on mannans by *R. intestinalis* and transcriptomic studies were performed by me.

The paper has been accepted for publication in Nature Communication (January 2019).

Supplementary information is included in the end of the chapter

The Human Gut Firmicute *Roseburia intestinalis* is a Primary Degrader of Dietary β-Mannans

Sabina Leanti La Rosa[1], Maria Louise Leth[2], Leszek Michalak[1], Morten Ejby Hansen[2], Nicholas A. Pudlo[3], Robert Glowacki[3], Gabriel Pereira[3], Christopher T. Workman[2], Magnus Ø. Arntzen[1], Phillip B. Pope[1], Eric C. Martens[3], Maher Abou Hachem[2], Bjørge Westereng[1,*]

Affiliations:

[1]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway.

[2]Department of Systems Biology, Danish Technical University, Lyngby, Denmark.

[3]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan, USA

*To whom correspondence should be addressed: bjorge.westereng@nmbu.no

Running title: β-Mannan Processing by *Roseburia intestinalis*

Key words: β-Mannan, *Roseburia intestinalis*, Short-Chain Fatty Acids, Carbohydrate Active Enzymes, Dietary Fibers, Gut Microbiota

Abstract

β-Mannans are plant cell wall polysaccharides that are commonly found in human diets. However, a mechanistic understanding into the key populations that degrade this glycan is absent, especially for the dominant Firmicutes phylum. Here, we show that the prominent butyrate-producing Firmicute *Roseburia intestinalis* expresses two loci conferring metabolism of β-mannans. We combine multi-"omic" analyses and detailed biochemical studies to comprehensively characterize loci-encoded proteins that are involved in β-mannan capturing, importation, de-branching and degradation into monosaccharides. In mixed cultures, *R. intestinalis* shares the available β-mannan with *Bacteroides ovatus*, demonstrating that the apparatus allows coexistence in a competitive environment. In murine experiments, β-mannan selectively promotes beneficial gut bacteria, exemplified by increased *R. intestinalis*, and reduction of mucus-degraders. Our findings highlight that *R. intestinalis* is a primary degrader of this dietary fiber and that this metabolic capacity could be exploited to selectively promote key members of the healthy microbiota using β-mannan-based therapeutic interventions.

Introduction

The human gastrointestinal tract harbors an extremely dense and diverse microbial community, known as the gut microbiota[1]. In a mutually beneficial relationship, the gut microbiota supplies enzymes able to depolymerize dietary carbohydrates that cannot be hydrolyzed by human enzymes[2,3]. The monosaccharides generated are further fermented into host-absorbable metabolites, including the short-chain fatty acids butyrate, acetate and propionate. In particular, butyrate produced by commensal bacteria serves as the main energy source for colonocytes[4,5] and it exhibits anti-carcinogenic, anti-inflammatory and barrier protective properties in the distal gut[6,7,8]. The relevance of this metabolic output to human health has prompted increasing interest in intentionally modulating the composition of the gut microbiota to promote wellbeing and combat disease, e.g. by the use of prebiotics[9]. Established prebiotics have been traditionally developed based on their selective fermentation by bifidobacteria and lactobacilli generally regarded as conferring health benefits to the host. Notably, other potentially beneficial targets are the butyrate-producing Firmicutes[9].

Roseburia spp., together with *Faecalibacterium prausnitzii* and *Eubacterium rectale,* constitute a group of dominant butyrate-producing Firmicutes, estimated to account for 7–24% of the total bacteria in the healthy human colon[10,11]. Interest in *Roseburia* spp. has increased with reports that the abundance of these bacteria is reduced in individuals affected by inflammatory diseases[12-14] and colorectal cancer[15,16]. Complementary studies have shown that *Roseburia* spp. play an important role in the control of gut inflammatory processes[17], amelioration of atherosclerosis[18] and in the maturation of the immune system, primarily through the production of butyrate[19]. *R. intestinalis* preferentially colonizes the mucin layer[20,21] and this intimate

association to the host may contribute to the local level of butyrate available for the colonic epithelial cells[22]. This species appears to be a specialist able to grow only on a few glycans[23,24] and has been recently shown to be a prominent xylan degrader *in vitro*[25] and in the healthy human colon[26].

β-Mannans are widespread in the human diet: they are widely used in food as thickening, stabilizing and gelling agents[27] (glucomannan and galactomannan, Fig.1). They are found in the endospermic tissue of nuts (homopolymeric mannan), coffee beans, coconut palm, tomato and legume seeds (galactomannan) (Fig.1)[27-29], and play vital roles in the cell wall structure and as storage polysaccharides in plants. Notably, the structure of galactoglucomannan[29] from non-edible sources (softwood) shares striking similarities with that from edible sources (Fig.1).

Prevalent Gram-negative *Bacteroides* spp. encode β-mannan polysaccharide utilization loci (PULs) and have been recently shown to utilize mannans[30-32]. Despite members of the Firmicutes phylum being numerically dominant in the gut, insight is lacking into the metabolic strategies adopted by these Gram-positive bacteria to utilize β-mannans.

Here, using a combination of microbiology,"omic" and enzymology approaches, we unravel the molecular mechanism evolved by *R. intestinalis* L1-82 to depolymerize β-mannans that are potentially available in the large intestine. Our findings show that *R. intestinalis* growth on β-mannan is dependent on the expression of a highly specific multi-modular cell attached endomannanase, an ATP-binding transporter and an intracellular enzyme cocktail through which linear and substituted manno-oligosaccharides are completely hydrolyzed to component monosaccharides for further metabolism. Using germ-free mice colonized with a model gut microbiota, we demonstrate that β-mannan alters the community composition, facilitating bacteria

85 that have mannan degrading machineries. Besides extending the knowledge on the
86 enzymatic basis of β-mannan-metabolism by members of the most numerous
87 Firmicutes phylum, our results have implications for the design of targeted
88 intervention strategies to manipulate the gut microbiota via supplementation of
89 prebiotics to the diet to restore or improve health.
90

RESULTS

Two multi-gene loci mediate β-mannan utilization. *R. intestinalis* L1-82 grows efficiently on a variety of complex β-mannans as a sole carbon source (Fig. 2a), causing a concomitant acidification of the medium (Fig. 2b). To evaluate which fractions of β-mannan breakdown products are internalized, we analyzed the culture supernatants during *R. intestinalis* growth on AcGGM using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (Supplementary Fig. 1a-b). Neither oligosaccharides nor monosaccharides accumulated in the stationary phase culture (Supplementary Fig. 1a-b), indicating that the bacterium possesses a highly efficient apparatus to cleave and import all the sugars derived from the breakdown of this complex glycan.

To examine the molecular basis underlying β-mannan utilization by *R. intestinalis*, we performed an RNA sequencing (RNAseq) transcriptional analysis during growth on konjac glucomannan (KGM), spruce acetylated galactoglucomannan (AcGGM) and glucose (Glc). The top 20 upregulated genes in β-mannan transcriptome encode a β-mannanase belonging to the glycoside hydrolase (GH) 26 family (GH26 according to the CAZy classification[33]), a solute binding protein (MnBP) and two permeases (MPP) of an ABC transporter, two phosphorylases (GH130), one epimerase (Mep), two β-glucosidases (GH3) and two carbohydrate esterases (CEs) (Fig. 2c and Supplementary Data 1). These genes are located in two loci, which were designated mannan-utilization locus large (MULL: ROSINTL182_05469-83) and mannan-utilization locus small (MULS: ROSINTL182_07683-85) (Fig. 2d). Among the MULL genes expression of a LacI-type transcriptional regulator, predicted glycosyl hydrolases belonging to GH113, GH36, GH1 and a phosphomutase also increased. The response was specific to β-mannan as no

differential expression of these genes was observed during growth of *R. intestinalis* on galactose, a building block in mannan (Supplementary Table 1).

Proteomic analysis under the same growth conditions corroborated the RNAseq results; indeed, proteins encoded by the genes located in MULL and MULS were abundant in the AcGGM samples compared to the glucose samples (Fig. 2e, Supplementary Data 2).

We carried out a comparative genomic analysis to establish the distribution of β-mannans utilization loci equivalent to the identified MULL and MULS in other representative Roseburia *spp*. and Clostridium cluster XIVa members. The results showed that *R. faecis and R. hominis* shared an overall MULL and MULS organization with that of *R. intestinalis* (Supplementary Fig. 2, Supplementary Table 2), suggesting that the utilization of β-mannan is shared by these three *Roseburia* spp. However, the lack of the critical GH26 endomannanase, required to break down mannan (see later results for *R. intestinalis* β-mannanase *Ri*GH26), is likely to render *R. hominis* only able to metabolize manno-oligosaccharides. Orthologous mannan utilization loci were identified in specific members of the Clostridium cluster XIVa, although a similar organization and complete conservation of all MULL and MULS components was not observed (Supplementary Fig. 2).

Degradation of the β-mannan backbone. *Ri*GH26, (locus tag: ROSINTL182_07683), is a predicted extracellular modular β-mannanase comprising a carbohydrate binding module of family 27 (CBM27), a catalytic module of GH26 followed by a CBM23 (Supplementary Fig. 3a). Furthermore, two C-terminal Ig-like domains and a proline-glycine rich region likely mediate cell attachment[34] and binding within the cell wall[35]. The extracellular localization of *Ri*GH26 was

corroborated experimentally by immunofluorescence microscopy (Fig. 3). *Ri*GH26 exhibited activity toward decorated mannans including KGM, carob galactomannan (CGM) and AcGGM (Fig. 4a and Supplementary Fig. 3b), generating linear and substituted manno-oligosaccharides. The enzyme was active on mannopentaose (M_5) and mannotetraose (M_4) but not mannobiose (M_2) (Supplementary Fig. 3c). Overall, the product profiles suggest capability of endo-action and indicates that *Ri*GH26 targets large polymers and can accommodate the galactose and acetyl decorations present in these substrates. Further analysis indicated that *Ri*GH26 is a potent enzyme as, when used at the concentration of 10 nM, it was able to hydrolyze high concentrations of spruce AcGGM (50 mg ml^{-1}) into oligosaccharides in 1 h at standard assay conditions (Supplementary Fig. 3d). No detectable activity was measured when *Ri*GH26 was incubated with linear cello-oligosaccharides, birch xylan, curdlan, lichenan or barley derived β-glucan, thus confirming the specificity of *Ri*GH26 towards β-mannan (Supplementary Fig. 3c).

BlastP searches showed that homologs of *Ri*GH26, including the two predicted carbohydrate binding modules CBM27 and CBM23, were exclusively found in β-mannanase encoded by Firmicutes belonging to various other members of the Clostridium cluster XIVa (Supplementary Fig. 4, Supplementary Table 3-5). To investigate the biochemical properties of the two modules, *Ri*CBM27 and *Ri*CBM23 were expressed in *Escherichia coli* and their capacities to bind to a range of different soluble cello- and manno-oligosaccharides were evaluated using surface plasmon resonance (SPR). Recombinant *Ri*CBM27 and *Ri*CBM23 bound only manno-oligosaccharides (Table 1), but differed in their binding profiles. Similar to a previously described GH26-associated CBM27[36], *Ri*CBM27 preferred mannohexaose (M_6) (K_d=165 ± 10 μM, two independent experiments, ± indicates

standard deviation), (Table 1, Supplementary Fig. 5a) and its affinity dropped for ligands smaller than a tretrasaccharide (Table 1). By contrast, *Ri*CBM23 was selective for shorter oligosaccharides with its highest affinity for M_4 (K_d=130 ± 50 µM, two independent experiments) (Table 1, Supplementary Fig. 5b), although mannotriose (M_3) was also bound with good affinity (Table 1).

Internalization of break-down products from β-mannan. Within the MULL cluster, the three genes (ROSINTL182_05477 – ROSINTL182_05479) that encode an ATP-binding cassette (ABC) importer were shown to exhibit the highest level of increased expression during growth on β-mannan (and when compared to growth on glucose). The thermodynamic binding parameters of the ABC-transporter associated solute binding protein, *Ri*MnBP, to linear and substituted manno-oligosaccharides were determined using isothermal titration calorimetry (ITC). *Ri*MnBP bound a range of unsubstituted manno-oligosaccharide with a preference for M_5 (K_d of 2.55 µM) followed by M_3 and M_4 (Table 2, Supplementary Fig. 6). Acetylations had a marginal effect on the binding affinities, thus providing evidence that these fragments are efficiently captured by the transport protein. Overall, these results support the predicted role of *Ri*MnBP in the uptake of manno-oligosaccharides generated by *Ri*GH26, showing optimal affinity for undecorated or acetyl substituted ligands with a degree of polymerization (DP) of 4−5.

Decomposition of internalized β-manno-oligosaccharides. The affinity of the solute binding protein *Ri*MnBP to manno-oligosaccharides and the predicted intracellular location of the debranching and exo-acting enzymes is consistent with a hierarchical degradation of the internalized manno-oligosaccharides following extracellular degradation of the β-mannan polymers by *Ri*GH26.

189 The ROSINTL182_05471 (*Ri*CEX) and ROSINTL182_05473 (*Ri*CE2) gene products
190 possess SGNH hydrolase-type esterase domain signatures[37]. Comparison to
191 previously characterized CEs revealed that *Ri*CE2 showed 25- 30% identity to a CE2
192 from *Clostridium thermocellum*[38] and the acetyl xylan esterase Axe2C of *Cellvibrio*
193 *japonicus*[38]. In contrast, *Ri*CEX did not display significant relatedness to other
194 characterized CE catalogued in the CAZy database[33], which excluded *Ri*CEX from
195 being classified in any of the 16 CE families. *Ri*CEX and *Ri*CE2 showed mannan
196 acetyl esterase activity on a mixture of oligosaccharides generated via *Ri*GH26
197 hydrolysis of AcGGM (Fig. 4b). *Ri*CE2 partially removed acetyl groups from the
198 acetylated oligosaccharide substrate (Fig. 4b). *Ri*CEX deacetylated the substrate
199 mainly to free and monoacetylated oligosaccharides (Fig. 4b). These results indicate
200 that *Ri*CEX has a preference for oligosaccharides with a degree of acetyl substitution
201 ≥ 2, but is less efficient on mono-substituted substrates. At the same time, it
202 suggests that an acetyl group present at a specific position (*O*-2 or *O*-3) is resistant
203 to enzymatic deacetylation by *Ri*CEX. The combination of *Ri*CEX and *Ri*CE2
204 resulted in the almost complete deacetylation of the manno-oligosaccharides,
205 indicating a cooperative interaction of the two esterases (Fig. 4b).

206 *Ri*GH36 released galactose from internally substituted CGM and AcGGM after the
207 treatment with the *Ri*GH26 β-mannanase (Fig. 4c and Supplementary Fig. 7).
208 Interestingly, *Ri*GH36 released galactose from CGM-endomannanase products with
209 100% efficiency (Fig. 4c, Supplementary Fig. 8a) as no oxidized product could be
210 observed after treatment of these samples with galactose oxidase. The enzyme
211 exhibited limited activity on large polymers (Supplementary Fig. 8b) consistent with
212 the activity on internalized oligosaccharides *in vivo*. Similarly, α-galactosidase activity
213 increased after de-acetylation of the oligosaccharides (Supplementary Fig. 8c-d).

214 Beside cleaving single internal galactose residues from manno-oligosaccharides, this
215 enzyme was capable of removing α-1,6-galactose from the reducing-end of a
216 substituted manno-oligosaccharide (Supplementary Fig. 8e) and from an
217 oligosaccharide containing two consecutive substitutions (Supplementary Fig. 8f).
218 Corroborating these results, *Ri*GH36 cleaved galactose decorations from
219 endomannanase products of highly substituted guar gum galactomannan
220 (Supplementary Fig. 8b).

221 Sequence searches showed that the protein encoded by ROSINTL182_05483
222 (MULL, *Ri*GH113) exhibited 40% identity to the only characterized enzyme from this
223 family, the endo-β-mannanase AaManA from *Alicyclobacillus acidocaldarius*[39]
224 (Supplementary Fig. 9a). Alignment of *Ri*GH113 and AaManA showed that the
225 catalytic and substrate interacting residues are conserved (Supplementary Fig. 9a).
226 When *Ri*GH113 was assayed for activity on linear manno-oligosaccharides, it
227 revealed an ability to cleave linear manno-oligosaccharides to yield mannose and M_2
228 (Supplementary Fig. 9b). Strikingly, time-course analysis of *Ri*GH113 activity
229 revealed that this enzyme displays a different sub-specificity by hydrolyzing manno-
230 oligosaccharides to mannose and a minor amount of M_2 (Supplementary Fig. 9c).
231 After overnight incubation with *Ri*GH113, M_2 was partially degraded to mannose
232 (Supplementary Fig. 9d), confirming the exo-mannosidase activity as opposed to the
233 endo-fashion cleavage reported for the AaManA. The substituted manno-
234 oligosaccharides galactosylmannobiose (Gal_1Man_2) and digalactosylmannopentaose
235 (Gal_2Man_5) were hydrolyzed to a lesser extent than non-substituted substrates
236 (Supplementary Fig. 9e); no activity could be detected on Gal_1Man_2 while Gal_2Man_5
237 was hydrolyzed to yield mannose and digalactosylmannotetraose (Gal_2Man_4), which
238 was resistant to further hydrolysis. When the reducing end of manno-

239 oligosaccharides was blocked (Supplementary Fig. 10a-d), no *Ri*GH113 activity
240 could be detected demonstrating that this enzyme possesses a previously unknown
241 reducing end mannose-releasing exo-oligomannosidase activity. Consistent with the
242 view that *Ri*GH113 is an intracellular enzyme, release of mannose was detected
243 after incubation of the enzyme with *Ri*GH26-generated CGM-oligosaccharides (Fig.
244 4d). The closest homologues of this enzyme are encoded by Clostridium cluster
245 XIVa strains and a range of Firmicutes (Supplementary Fig. 10e).

246 Product analysis of end point assays and a time course study revealed that both
247 *Ri*GH3A (ROSINTL182_07684) and *Ri*GH3B (ROSINTL182_07685) were β-
248 glucosidases, with redundancy in structure (Supplementary Fig. 11a-b), active on
249 linear cello-oligosaccharides (Fig. 4e). *Ri*GH3B completely hydrolyzed cellotetraose
250 (G_4) and cellopentaose (G_5) into glucose monomers, whereas *Ri*GH3A released
251 glucose and a range of oligosaccharides with lower efficiency compared to that of
252 *Ri*GH3B (Supplementary Fig. 11c-d). Neither of these enzymes were active on
253 manno-oligosaccharides (Supplementary Fig. 11e-f). While *Ri*GH3B was able to
254 hydrolyze glucosylmannose (G_1M_1) and, partially, mannosylglucose (M_1G_1) into
255 monomers (Supplementary Fig. 11f), *Ri*GH3A displayed activity only towards G_1M_1.
256 No activity was detected on polymeric KGM (Supplementary Fig. 12a), while glucose
257 was released after incubation of both *Ri*GH3A and *Ri*GH3B with *Ri*GH26-generated
258 KGM–hydrolysate (Supplementary Fig. 12b). Importantly, the latter results
259 demonstrate that *Ri*GH26 can accept a glucose moiety at the subsite +1, generating
260 oligosaccharides with a glucose residue at the non-reducing end.

261 Recombinant *Ri*GH130_2 (MULL, ROSINTL182_05474) phosphorolyzed M_4 into M_3,
262 M_2 and mannose-1-phosphate (M1P) while M_3 was processed to M_2 and M1P (Fig.
263 4f). The enzyme was inactive on cello-oligosaccharides (Supplementary Fig. 13).

264 *Ri*GH130_2 was functional only in the presence of inorganic phosphate, confirming
265 that *Ri*GH130_2 is a mannosyl-phosphorylase.

266 **Catabolism of mannobiose and mannosylglucose units.** The concerted action of
267 the MULL and MULS encoded enzymes described above on the oligosaccharides
268 generated by *Ri*GH26, suggest an intracellular accumulation of M_2. Hydrolysis of this
269 product into monosaccharides is accomplished through the action of two enzymes
270 encoded by the co-transcribed genes ROSINTL182_05476 (*Ri*Mep) and
271 ROSINTL182_05475 (*Ri*GH130_1).

272 *Ri*Mep was active on M_2 and cellobiose (G_2), releasing M_1G_1 and G_1M_1, respectively
273 (Fig. 5a). These data show that *Ri*Mep is an enzyme active on the reducing end
274 sugar and catalyzes the conversion of disaccharide substrates to the corresponding
275 C2 epimer. This enzyme exhibited epimerization activity not only for the substrate
276 but also for the product as, under high enzyme amount and long reaction time, it was
277 able to convert M_1G_1 and G_1M_1 to M_2 and G_2, respectively (Supplementary Fig. 14a).
278 In addition, *Ri*Mep exhibited epimerization activity towards oligosaccharides with a
279 DP greater than 2 but not on monosaccharides (Supplementary Fig. 14b).

280 ROSINTL182_05475 encodes a specific mannosylglucose phosphorylase belonging
281 to the GH130 subfamily 1[40]. *Ri*GH130_1 was inactive on G_1M_1 and oligosaccharides
282 with a DP ≥ 2 (Supplementary Fig. 15). *Ri*GH130_1 displayed activity only towards
283 M_1G_1 in the presence of inorganic phosphate, releasing glucose and M1P (Fig. 5b
284 and 5c).

285 **Catabolism of phosphorolysis products.** *Ri*Pgm catalyzes the interconversion of
286 M1P and mannose 6-phosphate (M6P) (Fig. 5d). In addition, the enzyme displayed
287 activity also against D-glucose 1-phosphate (G1P) yielding D-glucose 6-phosphate

(G6P) (Supplementary Fig. 16a), thus indicating that *Ri*Pgm is a phosphomannomutase (PMM) / phosphoglucomutase (PGM) which can use either glucose or mannose as substrate. Consistent with the presence of a predicted magnesium-binding loop (Supplementary Fig. 16b), the *Ri*Pgm-mediated catalytic activity was detected only when $MgCl_2$ was present in the reaction.

ROSINTL182_05469/70 encodes a predicted bi-functional protein consisting of an N-terminal glucosidase domain (*Ri*GH1_D1, aa 1-246) and a C-terminal family GH1 isomerase domain (*Ri*GH1_D2, aa 247-768). *Ri*GH1_D1 shares 44% identity to the previously characterized β-glucosidase *Tm*GH1 from *Thermotoga maritima*[41]. The recombinant *Ri*GH1_D1 displayed no catalytic activity against all the tested substrates, including G_4, M_4, M_5, M6P, G6P and fructose 6-phosphate (F6P). Thus, *Ri*GH1_D1 functional significance is currently unclear. *Ri*GH1_D2 is a phosphomannose isomerase catalyzing the interconversion of M6P into F6P (Fig. 5e).

***R. intestinalis* competes with *Bacteroides* for β-mannans.** The ability of *R. intestinalis* to capture, breakdown β-mannan and efficiently internalize manno-oligosaccharides may increase its fitness to compete with other resident β-mannan degraders, including the glycan generalist *Bacteroides* [30]. To test this hypothesis, we performed *in vitro* co-cultivation of *R. intestinalis* and the efficient β-mannan degrader *Bacteroides ovatus* ATCC 8483[30]. Both bacteria showed similar growth rates in individual cultures supplemented with AcGGM (Fig. 6a). Population estimates using qPCR indicated that, in the mixed cultures, both *B. ovatus* and *R. intestinalis* grew well during the exponential growth phase, suggesting that the bacteria shared the available carbon source and maintained coexistence. (Fig. 6b). During the stationary phase, when glycan availability is limited, the mean relative

abundance of *R. intestinalis* and *B. ovatus* in the culture was approximately 72.5% versus 27.5%, respectively. In contrast, *R. intestinalis* showed slow growth on mannose (Fig. 6c) and was outcompeted when co-cultured in this carbon source with *B. ovatus* (Fig. 6d).

***R. intestinalis* responds rapidly to β-mannan supplementation.** To elucidate whether dietary supplementation of β-mannan can result in expansion of key gut bacteria able to utilize this hemicellulose, germfree mice were colonized with a synthetic microbiota composed of 14 sequenced strains of human commensal gut bacteria[42]. Colonized mice were fed a high-fiber diet for 14 days before being switched to a series of diet regimes with a varying amount of AcGGM (Fig. 6e). Overall, the levels of four species (*R. intestinalis*, *Bacteroides uniformis*, *B. ovatus* and *Marvinbryantia formatexigens*) gradually increased at both AcGGM doses (Fig. 6f - i) and these strains were able to suppress the bacteria foraging on the glyco-protein rich mucus layer (*Akkermansia muciniphila*, *Bacteroides caccae*, *Bacteroides thetaiotamicron*, *Bacteroides intestinihominis*) (Fig. 6e and 6j-m) and the amino acids degraders (*E. coli*, *Clostridium symbiosum* and *Collinsella aerofaciens*) (Fig. 6e and 6n). Upon feeding of a fiber-deficient diet, the fecal bacterial abundance of the mucin-eroding bacteria, the sulfate-reducer *Desulfovibrio piger* and the three amino acid degraders (Fig. 6e) rapidly increased with a corresponding decline of the fiber-degrading species.

DISCUSSION

β-Mannans are widely present in the human diet as constituents of hemicellulose in beans, some nuts and food additives, but are recalcitrant to intestinal digestion by host enzymes. These glycans instead serve as a carbon source for mannanolytic

bacteria in the distal gastrointestinal tract, primarily Firmicutes and Bacteroidetes. Recent studies have characterized a few enzymes encoded by two polysaccharide utilization loci (PULs) implicated in the metabolism of galactomannan in *B. ovatus*[30,43] and homopolymeric mannan in *Bacteroides fragilis*[32]. To date, a full understanding of β-mannan utilization by Firmicutes, however, remains underexplored. The human gut butyrate-producing Firmicute *R. intestinalis* has previously been shown to utilize galactomannan and glucomannan as a carbon source[42] and possesses predicted genes for the utilization of these substrates[24]. However, no data are available relating the mannanolytic activity at a biochemical level. In this study, we show that two conserved loci, MULL and MULS, collectively provide *R. intestinalis* the capacity to depolymerize this plant polysaccharide. Detailed biochemical studies of the encoded enzymes allowed us to construct a model of sequential action for the mannan utilization system encoded by MULL-MULS (Fig. 7). The *Ri*GH26 and the mannan ABC uptake system components *Ri*MnBP/*Ri*MPP1/*Ri*MPP2 transcripts and proteins were the most upregulated in both the RNA sequencing and proteomic analyses, respectively (Fig. 2c and Fig. 2e). This highlights the crucial role of this endomannanase and the ABC transport system in the β-mannan metabolic pathway. *Ri*GH26 is the only enzyme displayed on the cell surface (Fig. 3), allowing direct access to the intact β-mannan polymers through dynamic capture mediated by two appended carbohydrate binding modules (*Ri*CBMs). The SPR data showed that *Ri*CBM23 displays approximately 7- and 5-fold higher affinity for M_3 and M_4, respectively, than *Ri*CBM27, suggesting that the two CBMs play different roles to mediate binding of *Ri*GH26 to mannans. The *Ri*CBMs' K_d values for the preferred manno-oligosaccharides were in the 100−200 μM range (Table 1). This moderate affinity to the bound substrate constitutes an

advantage as it has lower impact on the catalytic activity compared to canonical counterparts from other organisms, and suggests an evolutionary adaptation of *R. intestinalis* to compete with other microbial enzymes with canonical higher- affinity CBMs, but with reduced catalytic rates[44]. Reliance on multi-modular cell-wall anchored enzymes is a common feature in Firmicutes[45]; consistently, *Ri*GH26 organization was primarily found in β-mannanase from other *Roseburia* species and members of the Clostridium cluster XIVa (Supplementary Fig. 4, Supplementary Table 3-5). Multiplicity of CBMs provides a contrast with the system for mannan metabolism in *Bacteroides ovatus*[30,43], where the binding and catalytic activity are distributed between two surface located binding proteins and the single domain mannanase *Bo*Man26B.

Collectively, our results point to a model in which the smaller manno-oligosaccharides generated by *Ri*GH26 are imported through a dedicated β-mannan transport system consisting of *Ri*MnBP/*Ri*MPP1/*Ri*MPP2 (Fig. 7). In the cytoplasm, acetylated and galactosylated manno-oligosaccharides are systematically debranched by *Ri*CE2, *Ri*CEX and *Ri*GH36, and subsequently depolymerized. Removal of glucose units from glucomannan-oligosaccharides is carried out by *Ri*GH3A and *Ri*GH3B. Based on the highest transcriptional and protein regulation, the main depolymerization strategy for breakdown of unsubstituted manno-oligosaccharides is mediated by the activity of two synergistic mannoside phosphorylases (*Ri*GH130_2 and *Ri*GH130_1) and an epimerase (*Ri*Mep), similar to the mannan catabolic pathway proposed in the ruminal bacterium *Ruminococcus albus*[46]. A similar system has been reported in *B. fragilis*[32] and *B. ovatus*[30], although only composed of an epimerase and a mannosylglucose phosphorylase (GH130_1) that, together, process GH26s-generated M_2 units. The presence of the manno-

oligosaccharide phosphorylases *Ri*GH130_2 allows *R. intestinalis* to process undecorated manno-oligosaccharide of DP >2, consistent with the internalization of large manno-oligosaccharides generated by *Ri*GH26-hydrolysis of polymeric mannan. However, GH130_2s mainly catalyze the phosphorolysis of undecorated manno-oligosaccharides[47]. Removal of mannose units from substituted substrates is mediated by the reducing end mannose-releasing exo-oligomannosidase *Ri*GH113, which displays a previously undescribed specificity. The two different approaches based on the phosphorylases and the GH113 are likely to be a functional adaptation to accelerate the depolymerization process of mannan. Eventually, mannan catabolism fuels monosaccharide fermentation via glycolysis and leads to the production of butyrate, which is the primary energy source for host colonocytes[5,48]. Colonocytes oxidize butyrate to carbon dioxide[49], thereby keeping the epithelium hypoxic (<1% O_2). This condition promotes gut homeostasis by stabilizing the hypoxia-inducible transcription HIF that coordinates barrier protection in the mucosa[50,51]. Recently, it has been shown that antibiotic-mediated depletion of butyrate-producing Clostridia increases colonocytes oxygenation and drives aerobic pathogen expansion in the gut lumen, resulting in *Salmonella enterica*-induced gastroenteritis[52]. Importantly, *R. intestinalis* has been found to affect host histone epigenetic states, direct colonic epithelial cells metabolism away from glycolysis and towards fatty acid metabolism, reduce the levels of inflammatory markers and ameliorate atheriosclerosis in a diet-dependent fashion[18]. The athero-protective effect was in part attributed to butyrate, as this SCFA has been shown to inhibit key inflammatory pathways involved in cardiovascular disease development[18].

The absence of oligosaccharides from *R. intestinalis* AcGGM-spent supernatant (Supplementary Fig. 1a-b) demonstrates that the β-mannan degradation apparatus

412 is optimized for efficient uptake of all the products released by *Ri*GH26, maximize
413 intracellular breakdown and avoid nutrient leakage. This will limit the access to other
414 bacteria, such as *Bacteroides spp*., competing for the same resource. Using
415 AcGGM, we have shown that *R. intestinalis* and *B. ovatus*, which possesses an
416 equally complex β-mannan degrading system, shared the available resources and
417 maintained coexistence (Fig. 6b). Notably, *R. intestinalis* outcompeted *B. ovatus* in
418 the late exponential and stationary phase of growth; these results show that *R.*
419 *intestinalis* is capable to bind and import the remaining β-mannan breakdown
420 products (preferred by the *Ri*MnBP transport protein) more efficiently than *B. ovatus*.
421 Thus, it is likely that the β-mannan utilization apparatus provides *R. intestinalis* with a
422 selective advantage during nutrient limitation when microbial competition for the
423 available carbohydrates in the gut is intense. Understanding the mechanism by
424 which β-mannan is degraded by key commensal members of the gut is crucial to
425 designing intervention strategies through the use of targeted prebiotics which aim to
426 program or reprogram the composition of the microbiota to maximize human health.
427 Our *in vivo* study demonstrates that a diet supplemented with AcGGM can be used
428 to manipulate the gut microbiota and to facilitate the growth of species equipped with
429 a β-mannan degrading system, including *R. intestinalis* (Fig. 6e). This is supported
430 by the increase in the relative abundance of *R. intestinalis*, *B. uniformis* and *B.*
431 *ovatus*, which all possess enzymes able to degrade AcGGM (BACUNI_00371 -
432 BACUNI_00383; BACOVA_02087-02097 and BACOVA_03386-03406 respectively).
433 *R. intestinalis* was highly responsive to the AcGGM within a day, with a 10 to 30 fold
434 increase at the 2.5% and 7.5% AcGGM diet, corroborating its ability to respond
435 dynamically to variation in this dietary fiber. Intriguingly, *R. intestinalis*' response did
436 not last over the 7 day feeding treatment and the acetogen *M. formatexigens*

seemed to replace it. A cluster of genes with predicted functions in β-manno-oligosaccharide utilization (BRYFOR_07194- BRYFOR_07206) was identified in the genome of *M. formatexigens* (Supplementary Fig.17a). The results shown in Supplementary Fig. 17b-d suggest that *R. intestinalis* and *M. formatexigens* occupy different metabolic niches in the distal gut; the former consumes complex β-mannans, whereas the acetogen feasts on mono- and oligosaccharides. When in co-culture with either *R. intestinalis* or *B. ovatus*, *M. formatexigens* was outcompeted *in vitro* (Supplementary Fig. 17e-f). A previous study with gnotobiotic mice bi-associated with the prominent saccharolytic bacterium *B. thetaiotamicron* and *M. formatexigens* showed that the presence of *M. formatexigens* caused a decrease in the cecal levels of *B. thetaiotaomicron*, compared with mono-associated controls. Transcriptional and metabolic analyses demonstrated that *M. formatexigens* is capable of consuming a variety of plant-derived oligosaccharides and microbial and host-derived *N*-glycans (such as *N*-acetylglucosamine), suggesting that this ability could confer a fitness advantage when competing with the glycan-consuming *Bacteroides*[53]. Thus, it is likely that, when present as part of the synthetic microbial community described in this paper, *M. formatexigens* may be indirectly benefiting of either manno-oligosaccharides feeding/cross-feeding with other microorganisms or by its ability to grow mixotrophically, simultaneously utilizing organic carbon sources and formate or H_2 for energy[53]. Notably, *M. formatexigens* outcompeted *B. ovatus* at the 7.5% AcGGM diet, underscoring the competitiveness of this acetogen in a community setting. In the context of a complex microbial community, it is likely that *M. formatexigens* makes an important contribution to host nutrition improving fermentation by acting as a formate or H_2 sink and by generating acetate as main metabolic product [53].

462 Collectively, diet-induced changes involved the promotion of mannanolytic bacteria
463 producing propionate, acetate and butyrate, metabolites that are known to regulate
464 hepatic lipid, glucose homeostasis and health of the intestinal hepithelium[11]. These
465 SCFA-producers gained a competitive advantage over colonic mucin-degrading
466 bacteria. Given that intermittent dietary fiber-deprivation results in a thinner mucus
467 layer in mice, eventually enhancing pathogen susceptibility[42], our results support the
468 concept that β-mannan-based interventions not only could contribute to preventing
469 mucus barrier dysfunctions but also maintaining a gut environment that keeps
470 pathogenic bacteria away. If confirmed in humans, these findings may help to
471 prevent diseases affecting the integrity of the colonic mucus layer, such as ulcerative
472 colitis[54]. Indeed, the fact that the β-mannan degradation pathway is a core trait found
473 in the majority of the human gut microbiota[55] highlights the relevance of potential
474 therapeutic interventions through the use of β-mannan formulations to the general
475 population.

476 **METHODS**

477 **Glycans**. Carbohydrate substrates used in this study are listed in Supplementary
478 Table 6. All glycan stocks were prepared at 10 mg ml^{-1} in ddH_2O and sterilized by
479 filtration using a 0.22 µm membrane filter (Sarstedt AG & Co, Germany).
480 **Bacterial Strains and Growth Conditions**. Unless otherwise stated, *R. intestinalis*
481 L1-82[23] was routinely grown at 37 °C without agitation in an anaerobic cabinet
482 (Whitley A95 workstation, Don Whitley, UK) under an 85% N_2/ 10% H_2/ 5% CO_2
483 atmosphere. Growth experiments were carried out in YCFA medium (YCFA - Yeast
484 extract-Casein hydrolysate-Fatty Acids)[56] supplemented with 0.5 % (w/v) of the
485 specific carbohydrate to be examined. Overnight cultures (300 µl) were used to
486 inoculate 30 ml aliquots of YCFA plus the carbohydrate to be tested. These pre-
487 cultures were passaged at least three times on the same media to ensure cell growth
488 adaptation on a single carbon source prior to inoculation of the final cultures for
489 growth experiments, RNA-sequencing and proteomic analysis. Bacterial growth was
490 determined spectrophotochemically by monitoring changes in the optical density at
491 600 nm (OD_{600}). In addition, growth on turbid substrates was assessed by measuring
492 differences in pH compared to that of starting medium. Growth and pH curves are
493 averages of three biological replicates, with two technical replicates each. Routine
494 culturing of *Bacteroides ovatus* ATCC 8483 and *M. formatexigens* DSM 14469 was
495 in anaerobic Chopped Meat Medium[57] under static conditions at 37 °C.
496 **Transcriptomic Analysis by RNA Sequencing**. *R. intestinalis* was cultured in
497 triplicate on YCFA supplemented with 0.5% (w/v) glucose, galactose, KGM or
498 AcGGM as described above. Cells were harvested at mid-exponential phase and
499 RNA was extracted using the RNeasy Mini Kit (Quiagen) according to the
500 manufacturer's instructions. RNA-seq libraries were prepared using the ScriptSeq

Complete kit from Epicentre. Samples were paired-end sequenced on an Illumina Hiseq 4000 instrument at Beijing Genomics Institute (BGI). Analysis of the RNA-seq results was performed exactly as described in[25]. Differential gene expression analysis was performed with the DeSeq2 package[58].

Cloning, Overexpression, and Protein Purification. The genes encoding mature forms of the proteins described in this study were amplified from the *R. intestinalis* L1-82 genomic DNA (BioProject accession number PRJNA30005 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA30005]) by PCR, using appropriate primers (Supplementary Table 7). PCR products were generated using the Q5 High-Fidelity DNA Polymerase (New England BioLabs, United Kingdom) with 50 ng DNA as template. Prior to cloning the DNA fragment encoding *Ri*GH1_D2 (ROSINTL182_05469), sequence ambiguities at the 3'-end of ROSINTL182_05470 were corrected through sequencing the PCR product generated with the primers listed in the Supplementary Table 8. The gene ROSINTL182_07683 was synthesized without the N-terminal signal sequence predicted by SignalP v.4.1[59] (residues 1-27 from transcription start). The PCR amplicons were cloned into the pNIC-CH expression vector by ligation-independent cloning (LIC)[60]. The gene encoding *Ri*Mnbp (ROSINTL182_05479) was cloned in the vector pETM-11 following the method described elsewhere[25]. Recombinant proteins generally contained a C-terminal His_6-tag, although, in some cases, His-tag translation was prevented by the introduction of one or two stop codons at the end of the open-reading frame (*Ri*Mep, *Ri*GH36, *Ri*Pgm and *Ri*GH113). The His_6-tag was excluded to prevent interaction with putative C-terminal active or catalytic residues that could be detrimental to the enzymes' activity. Constructs were verified by sequencing (Eurofins, UK). Proteins were produced in *E. coli* BL21 Star (DE3) cells (Invitrogen) as previously

described[61]. Briefly, cells were cultured to mid-exponential phase in Tryptone Yeast extract (TYG) containing 50 mg ml^{-1} kanamycin at 25 °C. Protein overexpression was induced by adding isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 200 µM, followed by incubation for a further 16 h at 25 °C. Cells were harvested by centrifugation, sonicated and recombinant proteins were purified by either immobilized metal ion affinity chromatography (IMAC) or hydrophobic interaction chromatography (HIC). For IMAC purification, the clarified cell lysate was loaded onto 5 ml HisTrap HP Ni Sepharose columns (GE Healthcare) connected to an ÄKTA purifier FPLC system (GE Healthcare). Protein elution was achieved by using a linear gradient from 5 to 500 mM imidazole. *Ri*GH113, *Ri*GH36, *Ri*Mep and *Ri*Pgm were purified by HIC by loading the cell-free broth, adjusted to buffer A (1.5 M ammonium sulfate), onto a 5 ml HiTrap Phenyl FF (GE Healthcare) equilibrated with the same buffer. Protein was eluted by using a linear reverse gradient to 100 mM NaCl over 90 min at a flow rate of 2.5 ml min^{-1}. After IMAC and HIC, samples were concentrated and further purified by Size Exclusion Chromatography (SEC) using a HiLoad 16/60 Superdex G75 size exclusion column (GE Healthcare) and a running buffer consisting of 20 mM Tris-HCl pH 8.0 with 200 mM NaCl. Fractions containing the pure protein were combined, concentrated and buffer exchanged to 20 mM Tris pH 8.0, using a Vivaspin 20 (10-kDa molecular weight cutoff) centrifugal concentrators (Sartorius Stedim Biotech GmbH, Germany). Protein purity was estimated to be over 95% for all the enzymes using SDS-PAGE. Protein concentrations were determined using the Bradford assay (Bio-Rad, Germany).

Glycoside Hydrolase and Phosphorylase Activity Assays. Enzyme assays, unless otherwise stated, were carried out in 10 mM sodium phosphate buffer, pH 5.8, for up to 16 h at 37 °C and 700 rpm. Reactions with *Ri*GH130_1 and

551 *Ri*GH130_2 were prepared in 100 mM sodium phosphate buffer, pH 5.8. The activity
552 of *Ri*Pgm against M1P and G1P was tested in 10 mM sodium phosphate buffer, pH
553 5.8, supplemented with 1 mM $MgCl_2$. To determine the specificity of *Ri*GH113, the
554 recombinant protein was sequentially incubated with 0.1 mg ml^{-1} pre- reduced or
555 oxidized manno-oligosaccharides at 37 °C overnight. Reduction of manno-
556 oligosaccharides was conducted by incubating 1 mg ml^{-1} manno-oligosaccharides in
557 a volume of 75 µl with sodium borodeuteride ($NaBD_4$; 0.5 M in 100 mM NaOH)
558 solution. The reaction was incubated overnight at room temperature then quenched
559 by adding 25 µl of 25 mM sodium acetate. Oxidation of manno-oligosaccharides
560 reducing-end was obtained by incubating the substrates (1 mg ml^{-1}) withthe
561 *Neurospora crassa* cellobiose dehydrogenase (*Nc*CDH) overnight at 37 °C. Both
562 $NaBD_4$ and *Nc*CDH pretreated samples were diluted 10X in standard assay buffer
563 before addition of *Ri*GH113. Between three and five independent experiments were
564 performed to determine the enzyme activities.

565 **MALDI-TOF Mass Spectrometry of Reaction Products.** Reaction products
566 generated by the enzymes used in this study were analyzed by matrix-assisted laser
567 desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) as
568 described previously[62]. Briefly, 2 µl of a matrix, consisting of 9% 2,5-
569 dihydroxybenzoic acid (DHB) in 30% acetonitrile, were applied to an MTP 384
570 ground steel target plate TF (Bruker Daltonics, Germany). Sample (1 µl) was then
571 mixed with the matrix and dried under a stream of warm air. Samples were analyzed
572 with an Ultraflex MALDI-ToF/ToF instrument (Bruker Daltonics, Germany), equipped
573 with a Nitrogen 337 nm laser beam and operated in positive acquisition mode.
574 Results were analyzed using the Bruker FlexAnalysis software (version 3.3).

575 **HPAEC-PAD.** Mono- and oligosaccharides products were analyzed on a Dionex
576 ICS-3000 HPAEC system operated by the Chromeleon software version 7 (Dionex,
577 Thermo Scientific), as described previously[62]. Sugars were injected onto a CarboPac
578 PA1 2×250-mm analytical column (Dionex, Thermo Scientific) coupled to a
579 CarboPac PA1 2×50-mm guard column kept at 30 °C. Manno-oligosaccharides and
580 phosphorylated monosaccharides were eluted in 0.1 M NaOH at a flow rate of 0.25
581 ml min^{-1} by increasing the concentration of sodium acetate (NaOAc) exponentially
582 from 0 to 0.3 M over 26 min (from 9 to 35 min after injection), before column
583 reconditioning by 0.1 M NaOH for 10 min. Commercial manno-oligosaccharides with
584 DP2−6 were used as standards. For cello-oligosaccharides, the separation was done
585 using a multistep linear gradient going from 0.1 M NaOH to 0.1 M NaOH–0.1 M
586 NaOAc over 10 min, 0.1 M NaOH–0.14 M NaOAc after 14 min, 0.1 M NaOH–0.3 M
587 NaOAc at 16 min followed by a 2 min exponential gradient to 1 M NaOAc, before
588 reconditioning with 0.1 M NaOH for 9 min. Cello–oligosaccharides with DP 2–6 were
589 used as standards. For the analysis of disaccharides (G_1M_1 or M_1G_1) and
590 phosphorylated monosaccharides generated from the activity of *Ri*GH130_2, *Ri*Mep,
591 *Ri*GH130_1, *Ri*Pgm and *Ri*GH1, the elution was done at 0.25 ml min^{-1} using a 40
592 min program. The program started with 0.01 M potassium hydroxide (KOH) for 15
593 min, reaching the concentration of 0.1 M KOH at 25 min after injection and was kept
594 for additional 5 min at the same KOH concentration. Between each sample, the
595 column was re-equilibrated by running initial conditions for 10 min.
596 **Protein Cellular Localization.** Proteins of interest were detected using anti-sera
597 raised in rats (Eurogentec) against the corresponding recombinant *Ri*GH26 or the
598 previously characterized RiXyn10A[25].

For immunofluorescence microscopy, *R. intestinalis* cells were grown in YCFA containing 0.5% AcGGM, wheat arabinoxylan (WAX) or glucose to an OD_{600} of 0.8, collected by centrifugation (4,000 x *g* for 5 min) and washed twice in phosphate buffered saline (PBS). Cells were resuspended in 500 µl PBS and fixed by adding an equal volume of 2 X formalin (9% formaldehyde in PBS) on ice for 30 min. The bacterial pellet was washed twice with 1 ml PBS prior to resuspension in 1 ml of blocking buffer (1% bovine serum albumin, BSA, in PBS) and incubation at 4 °C for 16 h. After incubation the cell pellets were harvested by centrifugation and the supernatant discarded. For labelling, the bacteria were incubated with 0.5 ml of anti-sera (diluted 1:500 in blocking buffer) for 2 h at 25 °C. The cells were then pelleted, washed with 1 ml PBS and resuspended in 0.5 ml goat anti-rat IgG Alexa-Fluor 488 (Sigma-Aldrich), diluted 1:500 in blocking buffer and incubated 1 h at 25 °C. The cells were again harvested, washed with 1 ml PBS and suspended in 100 µl PBS containing one drop of ProLong Gold antifade reagent (Life Technologies). Labeled bacterial cells were mounted onto glass slides and secured with coverslips. Fluorescence microscopy was performed on a Zeiss AxioObserver equipped with the ZEN Blue software. Images were acquired using an ORCA-Flash4.0 V2 Digital CMOS camera (Hamamatsu Photonics) through a 100x phase-contrast objective. A HXP 120 Illuminator (Zeiss) was used as a fluorescence light source.

Analysis of the Bacterial Proteome.

R. intestinalis was grown in triplicate on YCFA supplemented with either 0.5 % (w/v) glucose or AcGGM, respectively, as a sole carbon source. Samples (10 ml) were harvested at the mid-exponential growth phase. Cell pellet was collected by centrifugation (4,500 x *g*, 10 min, 4 °C), resuspended in 50 mM Tris-HCl, 0.1% (v/v) Triton X-100, 200 mM NaCl, 1 mM dithiothreitol and disrupted by bead-beating using

three 60s cycles with a FastPrep24 (MP Biomedicals, CA). Proteins were precipitated with ice-cold trichloroacetic acid (TCA), final concentration of 10% (v/v), incubated on ice for 1 h, centrifuged (15,000 × *g*, 15 min, 4 °C) to pellet the precipitated proteins and washed with 300 µl ice-cold 0.01 M HCl in 90% acetone. Proteins were separated by SDS-PAGE with a 10% Mini-PROTEAN gel (Bio-Rad Laboratories, CA) and then stained with Coomassie brilliant blue R250. The gel was cut into five slices, after which proteins were reduced, alkylated, and in-gel digested according to a method published previously[63]. The peptides were dried under vacuum, solubilized in 0.1% (v/v) trifluoroacetic acid (TFA) and desalted using C_{18} ZipTips (Merck Millipore, Germany) according to the manufacturer's instructions. The peptide mixture from each fraction was analyzed using a nanoHPLC-MS/MS system as described previously[63], using a Q-Exactive hybrid quadrupole-orbitrap mass spectrometer (Thermo Scientific) equipped with a nano-electrospray ion source. Mass spectral data were acquired using Xcalibur (v.2.2 SP1).

MS raw files were processed with the MaxQuant software suite[64] (version 1.4.1.2) for identification and label-free quantification (LFQ) of proteins. Proteins were identified by searching MS and MS/MS data of peptides against the UniProtKB complete proteome of *R. intestinalis* L1-82 (4,698 sequences) supplemented with common contaminants (e.g. keratins, trypsin, and bovine serum albumin). In addition, reversed sequences of all protein entries were concatenated to the database for estimation of false-discovery rates (FDRs). Trypsin was set as proteolytic enzyme and two missed cleavages were allowed. Protein N-terminal acetylation, oxidation of methionines, deamidation of asparagines and glutamines and formation of pyro-glutamic acid at N-terminal glutamines were defined as variable modifications while carbamidomethylation of cysteines was used as a fixed modification. The "match

649 between runs" feature of MaxQuant, which enables identification transfer only
650 between samples from the same carbon source based on accurate mass and
651 retention time, was applied with default parameters. All identifications were filtered in
652 order to achieve a protein false discovery rate (FDR) of 1%. A protein was
653 considered "present" if it was detected in at least two of the three biological
654 replicates in at least one glycan substrate. Missing values were imputed from a
655 normal distribution (width of 0.3 and down shifted 1.8 standard deviations from the
656 original distribution) in total matrix mode and differential abundance analysis was
657 performed using an unpaired two-tailed Student's *t*-test with a permutation-based
658 FDR set to 0.05. Hierarchical clustering and heat map representations were
659 generated using Euclidean distance measure and average linkage using Perseus[65]
660 (version 1.5.5.3).
661 **Substrate Binding Assay using SPR.** The affinity of *Ri*CBM27 and *Ri*CBM23 to
662 soluble manno- and cello-oligosaccharides was evaluated by SPR using a Biacore
663 T100 (GE Healthcare). The two CBMs, diluted into 10 mM sodium acetate (pH 4.1)
664 to 2.3 µM, were immobilized on a NTA sensor chip (GE Healthcare) to a density of
665 3000−4000 response units (RU). Sensograms were recorded at 25 °C in
666 phosphate/citrate buffer (20 mM phosphate/citrate buffer; 150 mM NaCl; pH 6.5,
667 0.005% (v/v) P20 surfactant) at 30 µl per min with association and dissociation times
668 of 90 s and 240 s, respectively. CBMs binding was tested towards 0.2 nM – 1 mM of
669 carbohydrate ligands dissolved in the same buffer as above. Data were analyzed
670 using the Biacore T100 evaluation software, and equilibrium dissociation constants
671 (K_d) were obtained by fitting a single-site binding model to either the steady-state
672 response data or the full sensograms.

673 **ITC.** Binding of manno-oligosaccharides to *Ri*MnBP was measured at 25 °C in 10
674 mM sodium phosphate pH 6.5 using an ITC_{200} microcalorimeter (MicroCal). *Ri*MnBP
675 in the sample cell was titrated by 19 injections of carbohydrate ligand separated by
676 120 s. The following concentrations were used: 900 µM of M_3 in the syringe and
677 76.5 µM *Ri*MnBP in the sample cell; 1365 µM of M_4 or M_5 in the syringe and 91 µM
678 *Ri*MnBP in the sample cell; 2270 µM of M_6 in the syringe and 117 µM *Ri*MnBP in the
679 sample cell; 750 µM of diacetylated mannotetraose (M_4Ac_2) in the syringe and 50 µM
680 *Ri*MnBP in the cell; 1500 µM of diacetylated mannopentaose (M_5Ac_2) in the syringe
681 and 100 µM *Ri*MnBP in the cell. Thermodynamic binding parameters were
682 determined using the MicroCal Origin software (version 7.0).

683 **Competition experiments.** *R. intestinalis*, *B. ovatus* and *M. formatexigens* cells
684 were grown overnight under anaerobic conditions in YCFA containing 0.5% (w/v)
685 AcGGM (YCFA-AcGGM) as the sole carbon source. These subcultures were used to
686 inoculate, in approximately equal proportions (estimated by OD_{600}), 30 ml of the
687 same medium. A control culture of YCFA-AcGGM was also inoculated with either *R.*
688 *intestinalis*, *B. ovatus* or *M. formatexigens*. Growth (OD_{600}) was monitored for up to
689 24 h, withdrawing 1 ml samples for quantitative PCR (qPCR) analysis at selected
690 time points. Cells were pelleted, combined with 200 µl of TE buffer (pH 7.8) and
691 bead-beated for 2 min (FastPrep96, MP Biomedicals, CA) using ≤106 µm acid-
692 washed glass beads (Sigma-Aldrich). Genomic DNA was extracted using the Mag
693 Midi kit (LGC Group, UK) according to the manufacturer's instructions. qPCR was
694 performed in a LightCycler 480 II system (Roche, Germany) using specific primers
695 for each strain (Supplementary Table 9). In addition, a High-Resolution Melting
696 (HRM) analysis was conducted to evaluate the specificity of the amplification and the
697 lack of primer dimers. The raw data were imported into the LinReg PCR program[66]

698 and the calculated Cq values and PCR efficiency were used to deduce the ratio of *R.*
699 *intestinalis*, *B. ovatus and M. formatexigens* at each time point. Statistically
700 significant differences were determined using the unpaired two-tailed Student's *t*-
701 test.

702 **Human Gut Microbiota-associated Mice and Diets.** All experiments involving
703 animals complied with all relevant ethical regulations for animal testing and
704 research and were approved by the University of Michigan, University Committee for
705 the Use and Care of Animals. Germfree mice were colonized with a synthetic
706 microbiota composed of 14 fully sequenced human species according to the
707 methodology previously adopted by Desai et al[42]. Briefly, seven 6-week-old germfree
708 male wild-type Swiss Webster mice that had been raised on *ad libitum* access to a
709 high fiber chow diet (LabDiet 5013) and autoclaved distilled water were gavaged for
710 3 consecutive days with 200 µl each day of a mixture of the 14 different species.
711 Colonized mice were maintained on this high fiber diet for 14 days before being
712 switched to a series of diet regimes with varying fibers. This feeding sequence
713 consisted of 7 days of feeding on a gamma-irradiated Fiber-Free (FF) diet
714 (TD.140343, Harlan Teklad, USA) that does not contain AcGGM or related
715 molecules. Mice were then switched for 7 days to a custom version of the same diet
716 that contained AcGGM at 2.5% w/w, followed by a 7-day washout period on the FF
717 diet, and finally 7 days of feeding on a version of this diet containing AcGGM at 7.5%
718 w/w (in both AcGGM diets an equivalent amount of glucose was removed to
719 accommodate the prebiotic addition). Fecal samples were taken 1 day before and 1
720 day after each diet transition, effectively allowing us to measure changes in response
721 to AcGGM supplementation at 1 and 7 days post exposure relative to the FF diet.
722 The relative abundance of each microbial strain at sampled time points was

measured by qPCR, using previously validated primer sets, from total DNA extracted from freshly voided fecal pellets (stored at -20 °C until extraction) exactly as described previously[42]. Statistically significant differences were determined using the unpaired two-tailed Student's *t*-test.

Comparative Genomic Analysis. Identification of similar β-mannan catabolic genes in bacteria belonging to the Clostridium XIVa cluster was performed using the Gene Ortholog Neighbourhood viewer on the Integrated Microbial Genomes website (https://img.jgi.doe.gov). This was done using the genes encoding *Ri*GH26 (ROSINTL182_07683, GenBank ABYJ02000124.1:7167-11129) and *Ri*MEP (ROSINTL182_05476, GenBank ABYJ02000025.1:3200-4429) as the search homolog and the default threshold e-value of 1e-5. Then, a sequence comparison was conducted where each *R. intestinalis* L1-82 RefSeq annotated protein sequence was subjected to BLASTp searches against other Clostridium XIVa members. Sequences with coverage <60% and amino acid similarity <45% were excluded.

Data Availability. All data supporting the findings of this study are available within the article and Supplementary Information, or from the corresponding author upon request. The transcriptomic data described in this article are submitted under NCBI BioProject accession number PRJNA516396. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012448.

743 **REFERENCES**

744 1 Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I.
745 Host-bacterial mutualism in the human intestine. *Science* **307**, 1915-1920
746 (2005).
747 2 Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A.
748 Polysaccharide utilization by gut bacteria: potential for new insights from
749 genomic analysis. *Nat. Rev. Microbiol.* **6**, 121-131 (2008).
750 3 El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The
751 abundance and variety of carbohydrate-active enzymes in the human gut
752 microbiota. *Nat. Rev. Microbiol.* **11**, 497-504 (2013).
753 4 Hamer, H. M. et al. Review article: the role of butyrate on colonic function.
754 *Aliment. Pharmacol. Ther.* **27**, 104-119 (2008).
755 5 Roediger, W. E. Role of anaerobic bacteria in the metabolic welfare of the
756 colonic mucosa in man. *Gut* **21**, 793-798 (1980).
757 6 Wang, H. B., Wang, P. Y., Wang, X., Wan, Y. L. & Liu, Y. C. Butyrate
758 enhances intestinal epithelial barrier function via up-regulation of tight junction
759 protein Claudin-1 transcription. *Dig. Dis. Sci.* **57**, 3126-3135 (2012).
760 7 Fung, K. Y., Cosgrove, L., Lockett, T., Head, R. & Topping, D. L. A review of
761 the potential mechanisms for the lowering of colorectal oncogenesis by
762 butyrate. *Br. J. Nutr.* **108**, 820-831 (2012).
763 8 Canfora, E. E., Jocken, J. W. & Blaak, E. E. Short-chain fatty acids in control
764 of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* **11**, 577-591
765 (2015).
766 9 Gibson, G. R. et al. Expert consensus document: The International Scientific
767 Association for Probiotics and Prebiotics (ISAPP) consensus statement on the

768 definition and scope of prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 491-
769 502 (2017).
770 10 Duncan, S. H. et al. Proposal of *Roseburia faecis* sp. nov., *Roseburia hominis*
771 sp. nov. and *Roseburia inulinivorans* sp. nov., based on isolates from human
772 faeces. *Int. J. Syst. Evol. Microbiol.* **56**, 2437-2441 (2006).
773 11 Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human
774 colonic microbiota. *Environ. Microbiol.* **19**, 29-41 (2017).
775 12 Takahashi, K. et al. Reduced abundance of butyrate-producing bacteria
776 species in the fecal microbial community in Crohn's disease. *Digestion* **93**, 59-
777 65 (2016).
778 13 Chassard, C. et al. Functional dysbiosis within the gut microbiota of patients
779 with constipated-irritable bowel syndrome. *Aliment. Pharmacol. Ther.* **35**, 828-
780 838 (2012).
781 14 Kumari, R., Ahuja, V. & Paul, J. Fluctuations in butyrate-producing bacteria in
782 ulcerative colitis patients of North India. *World J. Gastroenterol.* **19**, 3404-
783 3414 (2013).
784 15 Wang, T. et al. Structural segregation of gut microbiota between colorectal
785 cancer patients and healthy volunteers. *ISME J.* **6**, 320-329 (2012).
786 16 Geng, J., Fan, H., Tang, X., Zhai, H. & Zhang, Z. Diversified pattern of the
787 human colorectal cancer microbiome. *Gut Pathog.* **5**, 2 (2013).
788 17 Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous
789 *Clostridium* species. *Science* **331**, 337-341 (2011).
790 18 Kasahara, K. et al. Interactions between *Roseburia intestinalis* and diet
791 modulate atherogenesis in a murine model. *Nat. Microbiol.* **3,** 1461-1471
792 (2018).

19 Jost, T., Lacroix, C., Braegger, C. & Chassard, C. Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. *Br. J. Nutr.* **110**, 1253-1262 (2013).

20 Van den Abbeele, P. et al. Arabinoxylans and inulin differentially modulate the mucosal and luminal gut microbiota and mucin-degradation in humanized rats. *Environ. Microbiol.* **13**, 2667-2680 (2011).

21 Van den Abbeele, P. et al. Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an *in vitro* gut model. *ISME J.* **7**, 949-961 (2013).

22 Nishino, K. et al. Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *J. Gastroenterol.* (2017).

23 Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. *Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52**, 1615-1620 (2002).

24 Sheridan, P. O., Martin J. C., Lawley, T. D. , Browne, H. P., Harris, H. M. B., Bernalier-Donadille, A., Duncan, S. H., O'Toole, P. W., Scott, K. P., Flint, H. J. Polysaccharide utilization loci and nutritional specialization in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genom.* **2** (2016).

25 Leth, M. L. et al. Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* **3**, 570-580 (2018).

816 26 Chassard, C., Goumy, V., Leclerc, M., Del'homme, C. & Bernalier-Donadille,
817 A. Characterization of the xylan-degrading microbial community from human
818 faeces. *FEMS Microbiol. Ecol.* **61**, 121-131 (2007).
819 27 Yamabhai, M., Sak-Ubol, S., Srila, W. & Haltrich, D. Mannan biotechnology:
820 from biofuels to health. *Crit. Rev. Biotechnol.* **36**, 32-42 (2016).
821 28 Schroder, R., Atkinson, R. G. & Redgwell, R. J. Re-interpreting the role of
822 endo-beta-mannanases as mannan endotransglycosylase/hydrolases in the
823 plant cell wall. *Ann. Bot.* **104**, 197-204 (2009).
824 29 Moreira, L. R. & Filho, E. X. An overview of mannan structure and mannan-
825 degrading enzyme systems. *Appl. Microbiol. Biotechnol.* **79**, 165-178 (2008).
826 30 Bagenholm, V. et al. Galactomannan catabolism conferred by a
827 polysaccharide utilization locus of *Bacteroides ovatus*: enzyme synergy and
828 crystal structure of a beta-mannanase. *J. Biol. Chem.* **292**, 229-243 (2017).
829 31 McNulty, N. P. et al. Effects of diet on resource utilization by a model human
830 gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with
831 an extensive glycobiome. *PLoS Biol.*
832 https://doi.org/10.1371/journal.pbio.1001637 (2013).
833 32 Kawaguchi, K. et al. The mannobiose-forming exo-mannanase involved in a
834 new mannan catabolic pathway in *Bacteroides fragilis*. *Arch. Microbiol.* **196**,
835 17-23 (2014).
836 33 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat,
837 B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids*
838 *Res.* **42**, D490-495 (2014).

839 34 Zhou, F., Chen, H. & Xu, Y. GASdb: a large-scale and comparative
840 exploration database of glycosyl hydrolysis systems. *BMC Microbiol.* **10**, 69
841 (2010).
842 35 Navarre, W. W. & Schneewind, O. Surface proteins of gram-positive bacteria
843 and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol.*
844 *Biol. Rev.* **63**, 174-229 (1999).
845 36 Roske, Y., Sunna, A., Pfeil, W. & Heinemann, U. High-resolution crystal
846 structures of *Caldicellulosiruptor* strain Rt8B.4 carbohydrate-binding module
847 CBM27-1 and its complex with mannohexaose. *J. Mol. Biol.* **340**, 543-554
848 (2004).
849 37 Akoh, C. C., Lee, G. C., Liaw, Y. C., Huang, T. H. & Shaw, J. F. GDSL family
850 of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534-552 (2004).
851 38 Montanier, C. et al. The active site of a carbohydrate esterase displays
852 divergent catalytic and noncatalytic binding functions. *PLoS Biol.*
853 https://doi.org/10.1371/journal.pbio.1000071 (2009).
854 39 Zhang, Y. et al. Biochemical and structural characterization of the intracellular
855 mannanase AaManA of *Alicyclobacillus acidocaldarius* reveals a novel
856 glycoside hydrolase family belonging to clan GH-A. *J. Biol. Chem.* **283**,
857 31551-31558 (2008).
858 40 Saburi, W. Functions, structures, and applications of cellobiose 2-epimerase
859 and glycoside hydrolase family 130 mannoside phosphorylases. *Biosci.*
860 *Biotechnol. Biochem.* **80**, 1294-1305 (2016).
861 41 Zechel, D. L. et al. Iminosugar glycosidase inhibitors: structural and
862 thermodynamic dissection of the binding of isofagomine and 1-

863 deoxynojirimycin to beta-glucosidases. *J. Am. Chem. Soc.* **125**, 14313-14323
864 (2003).
865 42 Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic
866 mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339-1353
867 (2016).
868 43 Reddy, S. K. et al. A beta-mannan utilization locus in *Bacteroides ovatus*
869 involves a GH36 alpha-galactosidase active on galactomannans. *FEBS Lett.*
870 **590**, 2106-2118 (2016).
871 44 Santos, C. R. et al. Dissecting structure-function-stability relationships of a
872 thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochem. J.*
873 **441**, 95-104 (2012).
874 45 Cockburn, D. W. & Koropatkin, N. M. Polysaccharide degradation by the
875 intestinal microbiota and its influence on human health and disease. *J. Mol.*
876 *Biol.* **428**, 3230-3252 (2016).
877 46 Kawahara, R. et al. Metabolic mechanism of mannan in a ruminal bacterium,
878 *Ruminococcus albus*, involving two mannoside phosphorylases and
879 cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase,
880 beta-1,4-mannooligosaccharide phosphorylase. *J. Biol. Chem.* **287**, 42389-
881 42399 (2012).
882 47 Ladeveze, S. et al. Role of glycoside phosphorylases in mannose foraging by
883 human gut bacteria. *J. Biol. Chem.* **288**, 32370-32383 (2013).
884 48 Bar, F. et al. Mitochondrial gene polymorphisms that protect mice from colitis.
885 *Gastroenterology* **145**, 1055-1063 (2013).

886 49 Donohoe, D. R. et al. The Warburg effect dictates the mechanism of butyrate-
887 mediated histone acetylation and cell proliferation. *Mol. Cell* **48**, 612-626
888 (2012).
889 50 Kelly, C. J. et al. Crosstalk between microbiota-derived short-chain fatty acids
890 and intestinal epithelial HIF augments tissue barrier function. *Cell Host*
891 *Microbe* **17**, 662-671 (2015).
892 51 Rivera-Chavez, F., Lopez, C. A. & Baumler, A. J. Oxygen as a driver of gut
893 dysbiosis. *Free Radic. Biol. Med.* **105**, 93-101 (2017).
894 52 Rivera-Chavez, F. et al. Depletion of butyrate-producing Clostridia from the
895 gut microbiota drives an aerobic luminal expansion of *Salmonella*. *Cell Host*
896 *Microbe* **19**, 443-454 (2016).
897 53 Rey, F. E. et al. Dissecting the *in vivo* metabolic potential of two human gut
898 acetogens. *J. Biol. Chem.* **285**, 22082-22090 (2010).
899 54 Johansson, M. E., Sjovall, H. & Hansson, G. C. The gastrointestinal mucus
900 system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 352-361
901 (2013).
902 55 Lloyd-Price, J. et al. Strains, functions and dynamics in the expanded Human
903 Microbiome Project. *Nature* **550**, 61-66 (2017).
904 56 Lopez-Siles, M. et al. Cultured representatives of two major phylogroups of
905 human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids,
906 and host-derived substrates for growth. *Appl. Environ. Microbiol.* **78**, 420-428
907 (2012).
908 57 Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B.
909 Bacteria of the human gut microbiome catabolize red seaweed glycans with

carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19786-19791 (2012).

58 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

59 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785-786 (2011).

60 Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069-6074 (1990).

61 Mackenzie, A. K. et al. A polysaccharide utilization locus from an uncultured bacteroidetes phylotype suggests ecological adaptation and substrate versatility. *Appl. Environ. Microbiol.* **81**, 187-195 (2015).

62 Agger, J. W. et al. Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc. Natl. Acad. Sci U.S.A.* **111**, 6287-6292 (2014).

63 Arntzen, M. O., Karlskas, I. L., Skaugen, M., Eijsink, V. G. & Mathiesen, G. Proteomic investigation of the response of *Enterococcus faecalis* V583 when cultivated in urine. *PLoS One* https://doi.org/10.1371/journal.pone.0126694 (2015).

64 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

65 Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731-740 (2016).

66 Ramakers, C., Ruijter, J. M., Deprez, R. H. & Moorman, A. F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett.* **339**, 62-66 (2003).

67 Varki, A. et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323-1324 (2015).

68 Louis, P. & Flint, H. J. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol. Lett.* **294**, 1-8 (2009).

Acknowledgements. The authors wish to thank M. Kjos (NMBU) for assistance with immunofluorescence microscopy. This work was supported by a grant from the Research Council of Norway (244259). Additional funding was from a Graduate School DTU Scholarship, Lyngby, Denmark for M.L.L and from the Independent Research Fund Denmark, Natural Sciences (DFF, FNU) by a Research Project 2 grant (Grant ID: 4002-00297B) to M.A.H. Carlsberg Foundation is also acknowledged for an instrument grant for purchase of the ITC_{200} calorimeter to M.A.H.

Author contributions. Experiments were primarily designed by S.L.L.R., E.C.M., M.A.H. and B.W. S.L.L.R. cloned, expressed, purified and performed functional characterizations of the enzymes. Production of AcGGM was performed by L.M. and B.W. The initial growth experiments on mannans were performed by M.L.L. and these experiments were used to prepare RNA and performed the transcriptional analysis together with C.T.W. Proteomic analysis was done by S.L.L.R and M.Ø.A. SPR and ITC was performed by M.E.H., who also cloned, expressed and purified the

transport protein. S.L.L.R and G.P. conducted the *in vitro* growth experiments, competition experiments and qPCR. S.L.L.R performed enzyme localization studies. Mice experiment was conducted by N.A.P., R.G. and E.C.M. The manuscript was written primarily by S.L.L.R and B.W. with contributions from P.B.P., M.Ø.A., M.A.H., E.C.M. Figures were prepared by S.L.L.R., M.E.H. and N.A.P. All authors reviewed the final manuscript.

Competing interests.

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. General structure of the main classes of β-mannan. Linear homopolymeric (upper structure) and linear heteropolymeric (lower three structures) β-mannans are shown. Monosaccharides (D-mannose, green circle; D-glucose, blue circle; D-galactose, yellow circle) and acetylations (2Ac, 2-*O*-Acetyl; 3Ac, 3-*O*-Acetyl; 6Ac, 6-*O*-Acetyl) are represented using the standard Consortium of Functional Glycomics symbols[67]. Abbreviations: NR end, non-reducing end; R end, reducing end.

Figure 2. *R. intestinalis* upregulates several glycoside hydrolases, two carbohydrate esterases and an ABC-transporter during β-mannan consumption. a, Growth curves of *R. intestinalis* in YCFA without carbon source (black) or supplemented with 0.5% of either glucose (Glc, green circles), KGM (blue circles), CGM (purple circles) or AcGGM (red circles). **b**, pH measurements during *R. intestinalis* growth on Glc and β-mannans. In panel **a** and **b**, each point on the curves represent the average of three independent experiments. Error bars represent standard deviations (s.d.). **c**, RNA expression profile of putative β-mannan utilization genes during *R. intestinalis* L1-82 growth in YCFA supplemented with 0.5% KGM (blue bars) or AcGGM (red bars). The Log_2-fold change relative to cells cultured on YCFA-Glc is shown on the y-axis while the x-axis shows the putative genes involved in β-mannan catabolism. **d**, Genomic organization of the large and small β-mannan utilization loci (MULL and MULS, respectively) from *R. intestinalis*. Genes with similar predicted functions are coded by the same color. **e**, Heat map showing the proteomic detection of relevant proteins with predicted β-mannan utilization functions in triplicate samples (1−3) grown on YCFA-Glc and YCFA-AcGGM. Colors represent protein intensity expressed as Log_2 of LFQ values; the

color gradient ranges from 14 (green) to 34 (magenta), with black indicating 24. In panels c, d and e, locus tag numbers ROSINTL182_XXXXX are abbreviated with the last numbers after the hyphen.

Figure 3. Cellular location of the endomannanase *Ri*GH26. a, Fluorescent microscopy images of *R. intestinalis* cells cultured on AcGGM or Glc and incubated with polyclonal antibodies raised against the recombinant endomannanase *Ri*GH26. Glucose-grown cells exhibit a low intensity fluorescence signal; this is consistent with the results of the proteomics data showing that, when the organism is cultured on glucose, *Ri*GH26 is expressed at basal levels. **b**, Fluorescent microscopy images of *R. intestinalis* cells grown on WAX (positive control) and incubated with antibodies raised against the known surface endoxylanase *Ri*Xyn10A[25]. Localization microscopy images are representative data from two biological duplicates.

Figure 4. Cleavage of the β-mannans backbone, removal of the side chains and further depolymerization of the resulting linear manno-oligosaccharides. a, HPAEC chromatograms showing the oligosaccharide products after overnight digestion of KGM, CGM and AcGGM with *Ri*GH26. Samples were analyzed with the following manno-oligosaccharides as external standards: M_2, mannobiose; M_3, mannotriose; M_4, mannotetraose; M_5, mannopentaose; M_6, mannohexaose. **b**, MALDI-TOF analysis of *Ri*GH26-digested AcGGM incubated with either *Ri*CE2, *Ri*CEX or both enzymes. Peaks are labeled by DP and number of acetyl groups (Ac). **c**, HPAEC chromatograms showing products generated from CGM pre-digested with *Ri*GH26 and subsequently treated with *Ri*GH36. Assignments for peaks not occurring in the standard samples are based on comparison with the product profiles obtained by MALDI-TOF MS of *Ri*GH26-digested CGM (black spectrum) treated with galactose oxidase (GOX; brown spectrum). GOX converts a galactose residue in the

oligosaccharides into its corresponding aldehyde, giving a mass-to-charge ratio (*m/z*) of -2. All assigned masses are sodium adducts. Abbreviations: Ox, oxidation; Gal_1, galactose; Gal_1M_3, galactosylmannotriose; Gal_1M_4, galactosylmannotetraose; Gal_2M_4, digalactosylmannotetraose; Gal_2M_5, digalactosylmannopentaose. **d**, Product profiles from *Ri*GH26-digested CGM degradation experiments with *Ri*GH113 analyzed by HPAEC-PAD. The release of mannose confirms the exo-activity of *Ri*GH113. **e**, HPAEC-PAD traces showing activity of *Ri*GH3A or *Ri*GH3B towards G_5 and G_4 with the corresponding controls (no enzyme). Product profiles at various time points during the reaction are shown in Supplementary Fig. 6. Taken together, the data show that *Ri*GH3B is able to hydrolyze completely both tetramers and pentamers, producing glucose. *Ri*GH3A shows exo-activity towards both substrates that are converted slowly to glucose and a mixture of cello-oligosaccharides. Samples were analyzed with the following cello-oligosaccharides as external standards: glucose, G_1; cellobiose, G_2; cellotriose, G_3; cellotetraose, G_4; cellopentaose, G_5. **f**, Chromatograms showing products generated upon incubation of *Ri*GH130_2 with M_4 and M_3. The M1P released (red arrow) was identified by co-migration with the appropriate standard. In all panels, the data displayed are representative of at least three biological triplicates.

Figure 5. Enzymes for catabolism of mannobiose, mannosylglucose and monosaccharides deriving from complex β-mannan degradation. a, HPAEC-PAD traces showing the epimerization of M_2 and G_2 by *Ri*Mep to release M_1G_1 and G_1 M_1**,** respectively. **b-c**, HPAEC-PAD of phosphorolysis reactions of *Ri*GH130_1 using premixed M_1G_1 plus M_2 from Megazyme. The reaction products were then analyzed using an HPAEC method designed for the separation of **b**, phosphorylated manno-oligosaccharides or **c**, mono- and di-saccharides. M1P, indicated with a red

1043 arrow) and G_1 peaks in panel b are marked according to the standards **d**, HPAEC-
1044 PAD analysis of *Ri*Pgm-catalyzed conversion of M1P M6P. The M6P released was
1045 identified by co-migration with the M6P standard. **e**, Activity of *Ri*GH1_D2 on M6P
1046 and F6P analyzed by HPAEC-PAD.

1047 **Figure 6. *R. intestinalis* and *B. ovatus* co-culture experiments and *in vivo***
1048 **modulation of a synthetic human gut microbiota via AcGGM. a, c**, Growth rates
1049 of mono- and mixed cultures of *R. intestinalis* L1-82 (*Ri*) and *B. ovatus* ATCC 8483
1050 (*Bo*) on either AcGGM or mannose. Growth rate is defined as the maximum increase
1051 in absorbance at 600 nm (OD_{max}) divided by the time (T_{max}, in hours) to reach the
1052 maximum growth. **b, d**, *In vitro* competition experiment with *R. intestinalis* L1-82 and
1053 *B. ovatus* ATCC 8483 on either AcGGM or mannose as sole carbon source. The pH
1054 of the stationary phase cultures after growth on either AcGGM or mannose was 5.8 ±
1055 0.16 and 5.6 ± 0.11 (two biological triplicates, ± indicates the s.d.), respectively, thus
1056 showing that the results are not due to differences in acid sensitivity between the two
1057 strains. The relative abundance of the bacteria for each different phases of growth
1058 was determined by quantitative PCR of species-specific *vs* universal primers
1059 targeting the 16S rRNA genes. In panels a-d, the histogram bars show the mean of
1060 two biological replicates, with three independent measurements per replicate. Error
1061 bars represent s.d. Abbreviations: Early exp, early exponential phase; Middle Exp;
1062 middle exponential phase; Late exp, late exponential phase; Stat, stationary phase.
1063 **e**, Relative abundance of bacteria in fecal samples from germfree mice colonized
1064 with a synthetic human microbiota. Mice were shifted from a fiber-free (FF) diet to
1065 varying amounts (2.5% and 7.5% w/w) of AcGGM over time. Data are average of
1066 seven mice. **f-i**, Relative abundance of individual β-mannan-degrading bacteria and
1067 **j-m**, mucus-degraders**. n**, Additive relative abundances of three amino acids

1068 degraders. In panels f-n, histogram bars show the average of seven biological
1069 replicates while error bars represent s.d. *P*-values were calculated by two-tailed
1070 Student's *t* test. An asterisk (*) indicate a statistically significant difference ($P < 0.05$)
1071 in the relative abundance of each bacterium compared to that of the specific pre-FF
1072 diet. ns, not significant ($P \geq 0.05$).

1073 **Figure 7. Model for the degradation and utilization of complex β-mannans in *R.***
1074 ***intestinalis*.** Intracellular degradation of an acetylated galactoglucomanno-
1075 oligosaccharide is used as an example. Sugars are represented as in Fig. 1. Initial
1076 depolymerization of acetylated galactoglucomannan (AcGalactoglucomannan)
1077 occurs at the outer surface of *R. intestinalis* by the activity of *Ri*GH26 (green). The
1078 extracellular recruitment of β-mannan is facilitated by interactions with CBM27 and
1079 CBM23. Import of products occurs through the ABC transporter
1080 *Ri*MnBP/RiMPP1/RiMPP2 (orange). Within the cytoplasm, the acetyl and galactosyl
1081 decorations are removed by the two acetyl esterases *Ri*CE2 and *Ri*CEX (pink) and
1082 the α-galactosidase *Ri*GH36 (yellow). The two β-glucosidases *Ri*GH3A and *Ri*GH3B
1083 (blue) release glucose from the non-reducing end of the β-manno-oligosaccharide. In
1084 addition, the reducing end mannose-releasing exo-oligomannosidase *Ri*GH113
1085 (green) can catalyze the removal of mannose units from the decorated manno-
1086 oligosaccharides until it reaches a galactosyl substituent at the subsite -1. Once de-
1087 ornamented, the β-manno-oligosaccharides are saccharified by the exo-acting
1088 *Ri*GH130_2 (light green) with accumulation of M_2. The M_2undergoes subsequent
1089 epimerization and phosphorolysis by the concerted activity of *Ri*Mep - *Ri*GH130_1
1090 (light green), with release of glucose and M1P. These end products enter the
1091 glycolytic pathway either directly (for glucose) or after being converted into M6P and
1092 F6P by the phosphomannose mutase *Ri*Pgm (red) and the isomerase *Ri*GH1_D2

1093 (turquoise, purple domain). Released mannose is converted to M6P by a hexokinase
1094 and processed as described above. Galactose enters glycolysis after conversion to
1095 G1P via the Leloir pathway. The pyruvate generated from glycolysis is converted to
1096 acetyl-CoA and then butyrate. Black arrows show reactions demonstrated in this
1097 study. Green arrows indicate previously demonstrated steps for the generation of
1098 butyrate from monosaccharides fermentation[68] by *R. intestinalis*.

Table 1. Binding parameters of *Ri*CBM27 and *Ri*CBM23 to manno- and cello-oligosaccharides[a].

Ligand	K_d (μM)	
	*Ri*CBM27	*Ri*CBM23
M_3	1593 ± 30	230 ± 20
M_4	658 ± 20	130 ± 50
M_5	321 ± 20	198 ± 70
M_6	165 ± 10	205 ± 40
Glc_4	No binding	No binding
Glc_6	No binding	No binding

[a]Binding was determined by SPR. Values show the means and standard deviations of at least two independent experiments. K_d, dissociation constant.

Table 2. Thermodynamic binding parameters of *Ri*MnBP to linear and decorated manno-oligosaccharides[a].

Ligand	K_d (μM)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	-TΔS (kcal mol^{-1})	*n*
M_3	2.62	-7.6	-33.2	25.6	0.7
M_4	3.89	-7.4	-28.6	21.2	0.7
M_5	2.55	-7.7	-21.8	14.1	0.8
M_6	33.75	-6.2	-17.8	11.6	0.5
M_4Ac_2	25.65	-6.3	-21.9	15.6	0.9
M_5Ac_2	23.53	-6.3	-20.2	13.9	0.8

[a]Binding was measured by ITC. Data are means of two independent titrations. K_d, dissociation constant; ΔG, Gibbs free energy; ΔH, enthalpy; -TΔS, entropy; *n*, binding stoichiometry.

Figure 1

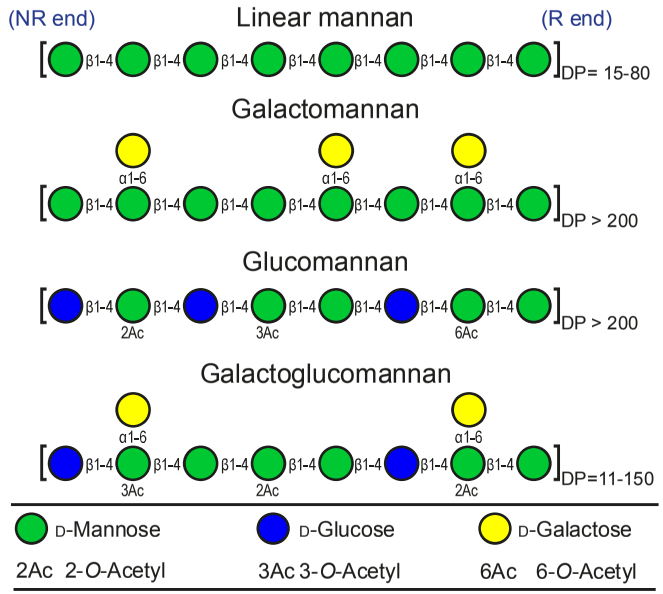

(NR end)
Linear mannan
(R end)
β1-4
DP= 15-80
Galactomannan
α1-6
DP > 200
Glucomannan
2Ac
3Ac
6Ac
Galactoglucomannan
DP=11-150
D-Mannose
D-Glucose
D-Galactose
2Ac 2-O-Acetyl
3Ac 3-O-Acetyl
6Ac 6-O-Acetyl

Figure 2

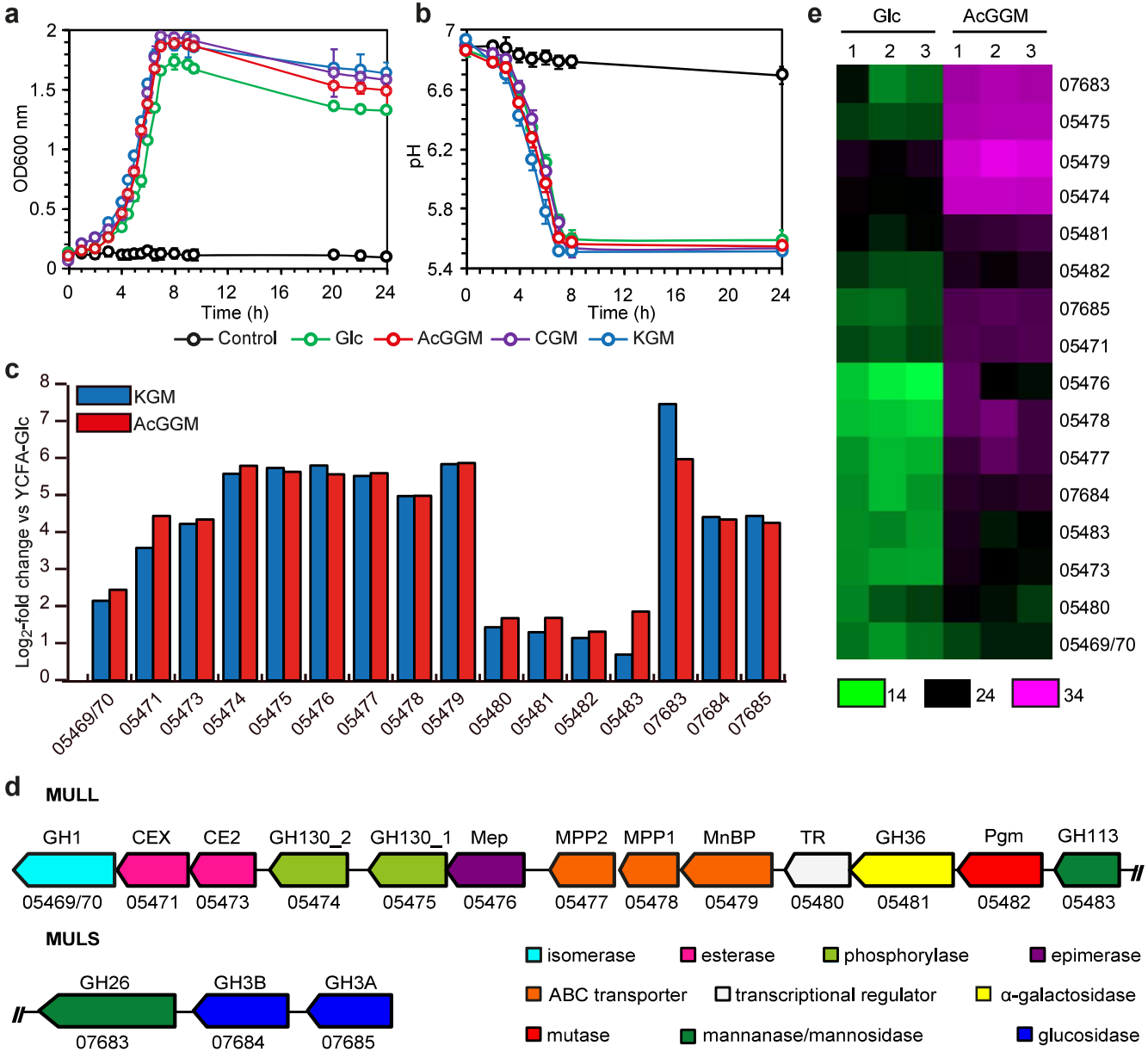

a
OD600 nm
Time (h)
b
pH
Time (h)
Control
Glc
AcGGM
CGM
KGM
c
KGM
AcGGM
Log2-fold change vs YCFA-Glc
05469/70
05471
05473
05474
05475
05476
05477
05478
05479
05480
05481
05482
05483
07683
07684
07685
e
Glc
AcGGM
1 2 3 1 2 3
07683
05475
05479
05474
05481
05482
07685
05471
05476
05478
05477
07684
05483
05473
05480
05469/70
14
24
34
d
MULL
GH1
CEX
CE2
GH130_2
GH130_1
Mep
MPP2
MPP1
MnBP
TR
GH36
Pgm
GH113
05469/70
05471
05473
05474
05475
05476
05477
05478
05479
05480
05481
05482
05483
MULS
GH26
GH3B
GH3A
07683
07684
07685
isomerase
esterase
phosphorylase
epimerase
ABC transporter
transcriptional regulator
α-galactosidase
mutase
mannanase/mannosidase
glucosidase

Figure 3

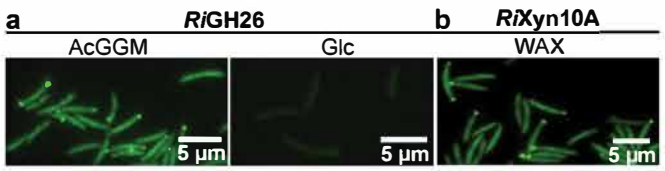


Figure 4

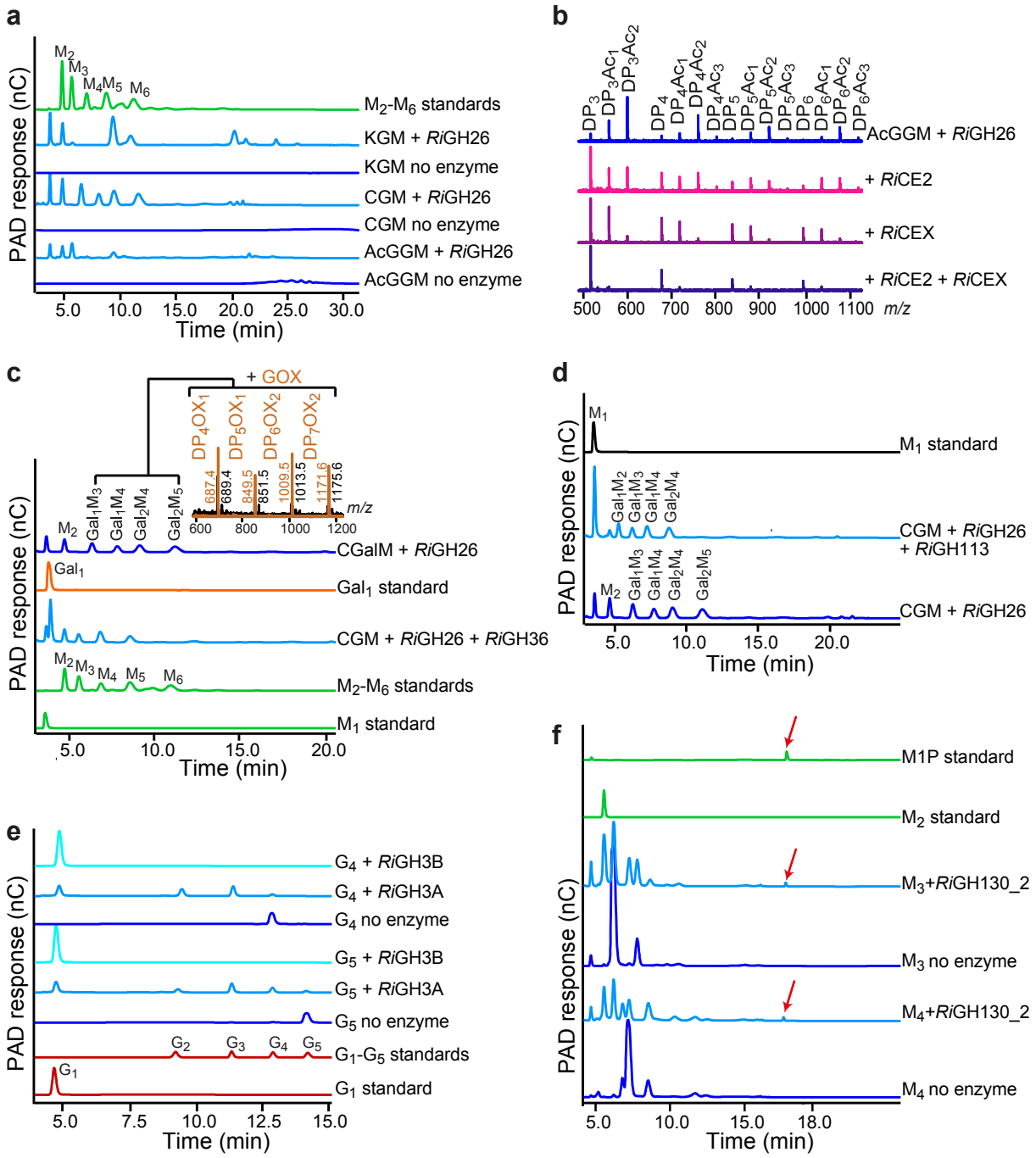

a
PAD response (nC)
Time (min)
M2-M6 standards
KGM + RiGH26
KGM no enzyme
CGM + RiGH26
CGM no enzyme
AcGGM + RiGH26
AcGGM no enzyme
b
AcGGM + RiGH26
+ RiCE2
+ RiCEX
+ RiCE2 + RiCEX
m/z
c
+ GOX
CGalM + RiGH26
Gal1 standard
CGM + RiGH26 + RiGH36
M2-M6 standards
M1 standard
d
M1 standard
CGM + RiGH26 + RiGH113
CGM + RiGH26
e
G4 + RiGH3B
G4 + RiGH3A
G4 no enzyme
G5 + RiGH3B
G5 + RiGH3A
G5 no enzyme
G1-G5 standards
G1 standard
f
M1P standard
M2 standard
M3+RiGH130_2
M3 no enzyme
M4+RiGH130_2
M4 no enzyme

Figure 5

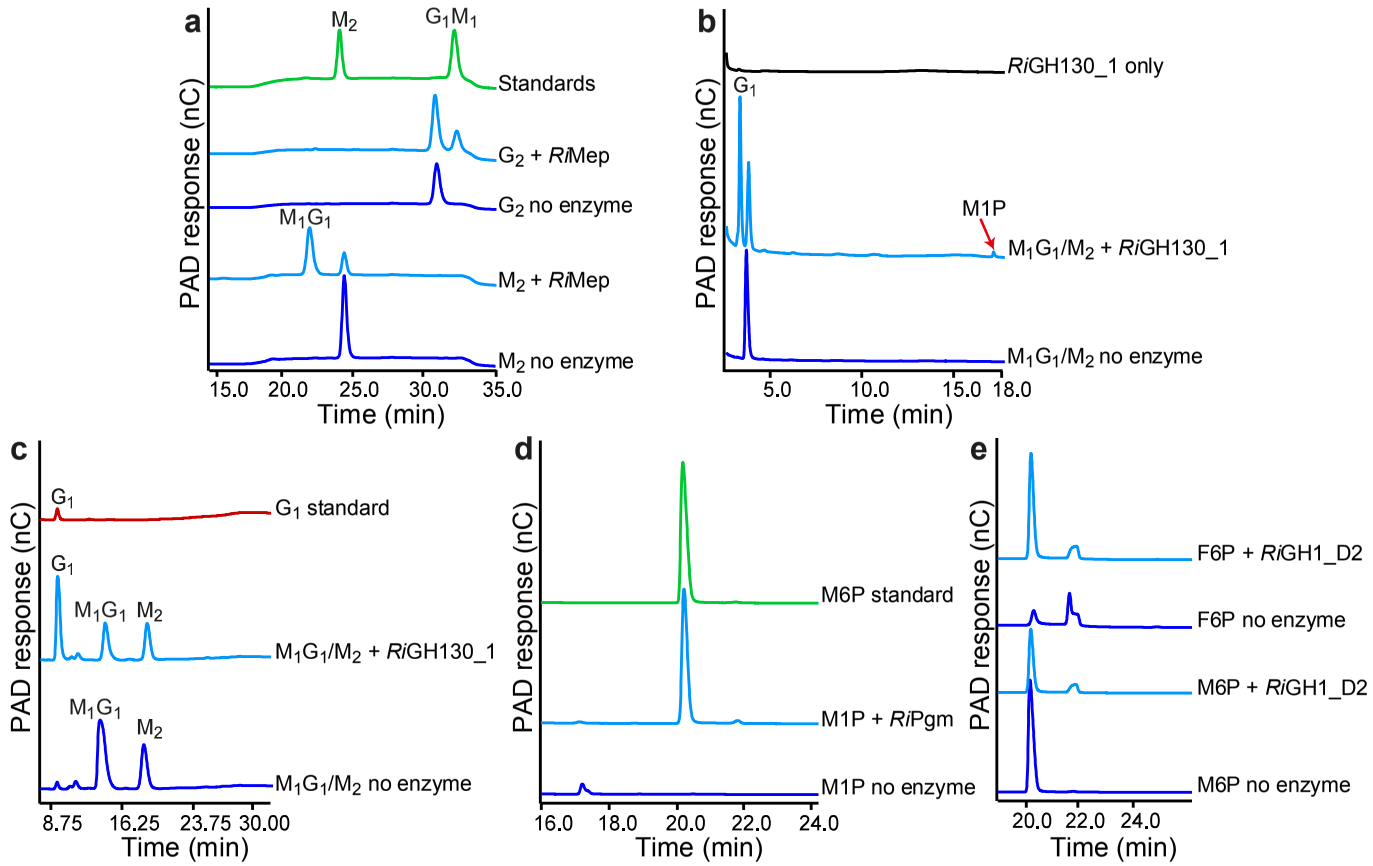

a
M2
G1M1
Standards
G2 + RiMep
G2 no enzyme
M1G1
M2 + RiMep
M2 no enzyme
PAD response (nC)
15.0 20.0 25.0 30.0 35.0
Time (min)
b
RiGH130_1 only
G1
M1P
M1G1/M2 + RiGH130_1
M1G1/M2 no enzyme
PAD response (nC)
5.0 10.0 15.0 18.0
Time (min)
c
G1
G1 standard
G1
M1G1
M2
M1G1/M2 + RiGH130_1
M1G1
M2
M1G1/M2 no enzyme
PAD response (nC)
8.75 16.25 23.75 30.00
Time (min)
d
M6P standard
M1P + RiPgm
M1P no enzyme
PAD response (nC)
16.0 18.0 20.0 22.0 24.0
Time (min)
e
F6P + RiGH1_D2
F6P no enzyme
M6P + RiGH1_D2
M6P no enzyme
PAD response (nC)
20.0 22.0 24.0
Time (min)

Figure 6

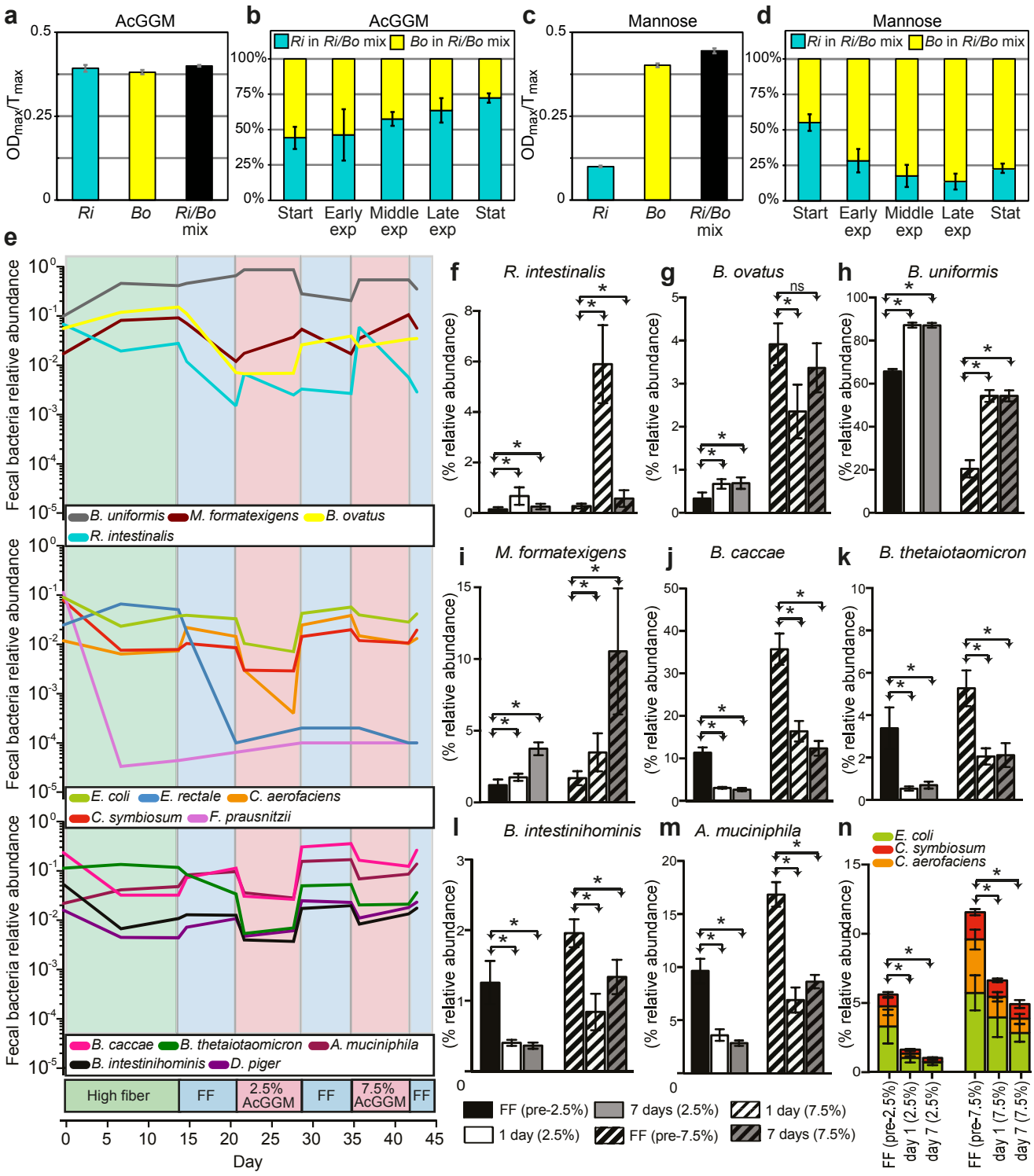
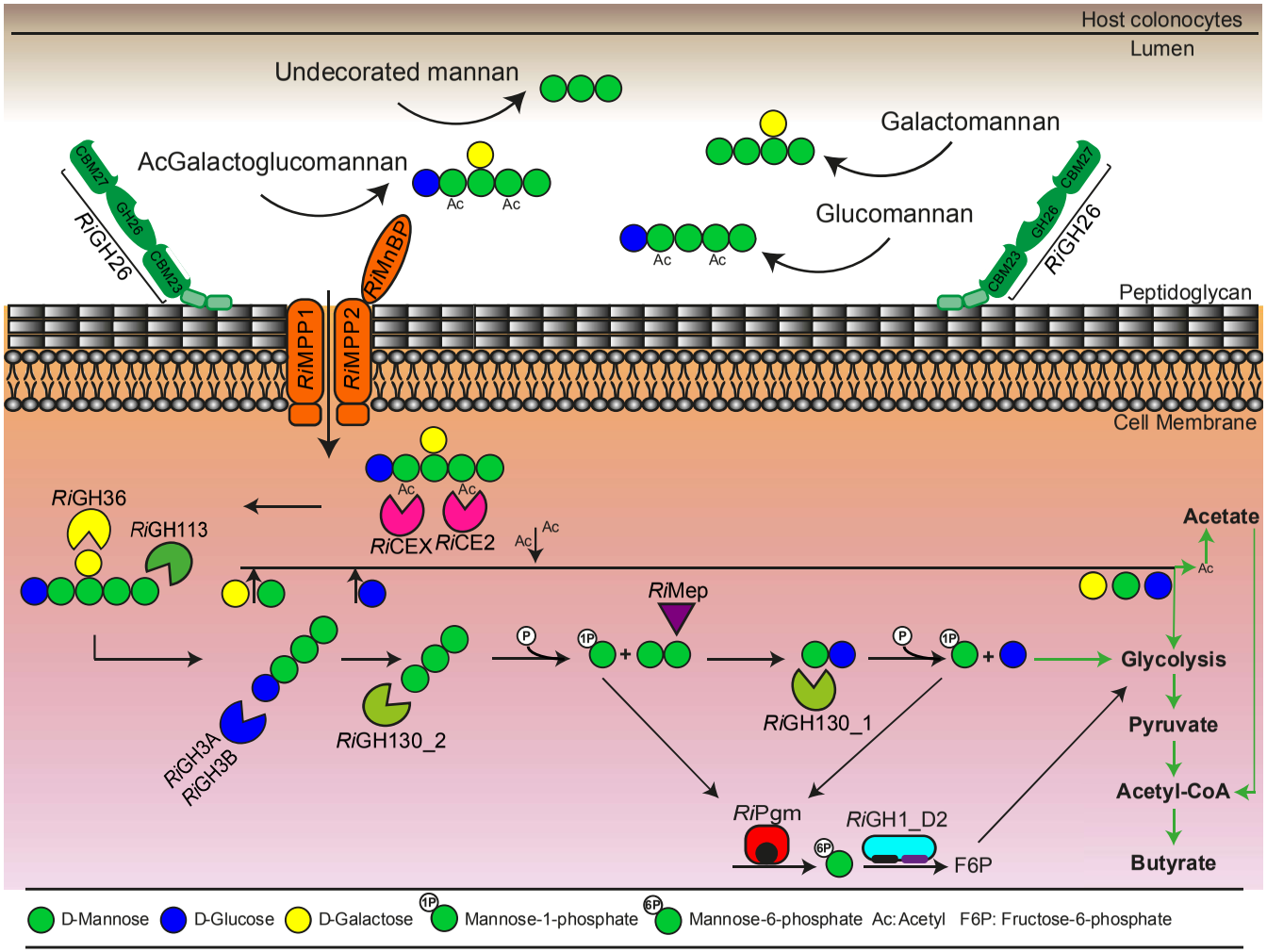

a
AcGGM
OD$_{max}$/T$_{max}$
Ri
Bo
Ri/Bo mix
b
AcGGM
Ri in Ri/Bo mix
Bo in Ri/Bo mix
Start
Early exp
Middle exp
Late exp
Stat
c
Mannose
d
Mannose
e
Fecal bacteria relative abundance
B. uniformis
M. formatexigens
B. ovatus
R. intestinalis
E. coli
E. rectale
C. aerofaciens
C. symbiosum
F. prausnitzii
B. caccae
B. thetaiotaomicron
A. muciniphila
B. intestinihominis
D. piger
High fiber
FF
2.5% AcGGM
FF
7.5% AcGGM
FF
Day
f
R. intestinalis
g
B. ovatus
ns
h
B. uniformis
(% relative abundance)
i
M. formatexigens
j
B. caccae
k
B. thetaiotaomicron
l
B. intestinihominis
m
A. muciniphila
n
E. coli
C. symbiosum
C. aerofaciens
FF (pre-2.5%)
7 days (2.5%)
1 day (7.5%)
1 day (2.5%)
FF (pre-7.5%)
7 days (7.5%)
FF (pre-2.5%)
day 1 (2.5%)
day 7 (2.5%)
FF (pre-7.5%)
day 1 (7.5%)
day 7 (7.5%)

Figure 7


Host colonocytes
Lumen
Undecorated mannan
Galactomannan
AcGalactoglucomannan
Glucomannan
Ac
CBM27
GH26
CBM23
RiGH26
RiMnBP
RiMPP1
RiMPP2
Peptidoglycan
Cell Membrane
RiGH36
RiGH113
RiCEX
RiCE2
Acetate
RiMep
Glycolysis
Pyruvate
Acetyl-CoA
Butyrate
RiGH3A
RiGH3B
RiGH130_2
RiGH130_1
RiPgm
RiGH1_D2
F6P
D-Mannose
D-Glucose
D-Galactose
Mannose-1-phosphate
Mannose-6-phosphate
Ac: Acetyl
F6P: Fructose-6-phosphate

Supplementary Information for

The Human Gut Firmicute *Roseburia intestinalis* is a Primary Degrader of Dietary β-Mannans

La Rosa et al.

Supplementary Table 1. Upregulation of MULL and MULS genes was not detected when *R. intestinalis* L1-82 grew on galactose.

Locus tag	Log_2-fold[a]
ROSINTL182_07683	-0.11
ROSINTL182_07684	-0.03
ROSINTL182_07685	0.03
ROSINTL182_05469	0.43
ROSINTL182_05470	-0.21
ROSINTL182_05471	-0.24
ROSINTL182_05473	-0.27
ROSINTL182_05474	-0.38
ROSINTL182_05475	-0.78
ROSINTL182_05476	-1.01
ROSINTL182_05477	-0.49
ROSINTL182_05478	-1.38
ROSINTL182_05479	-0.63
ROSINTL182_05480	-0.20
ROSINTL182_05481	-0.01
ROSINTL182_05482	-0.11
ROSINTL182_05483	-0.95

[a]Log_2-fold changes of the putative β-mannan utilization genes (as shown in Fig. 1d) expressed by *R. intestinalis* cells grown on galactose relative to glucose.

Supplementary Table 2. Sequence similarity between *R. intestinalis* L1-82's proteins involved in β-mannan utilization and other *Roseburia* strains/species.

		R. intestinalis M50/1	*R. intestinalis* XB6B4	*R. faecis* M72/1	*R. hominis* A2-183	*R. inulinovorans* DSM16841	*R. cecicola* GM
Locus Tag	**Function**	**Amino Acids Similarity (%) to *R. intestinalis* L1-82**					
05470/69	GH1	99	99	83	92	90	0
05471	CEX	99	99	75	76	0	0
05473	CE2	99	98	76	75	0	0
05474	GH130_2	99	100	85	95	0	0
05475	GH130_1	100	100	95	98	0	0
05476	Mep	99	99	85	89	0	0
05477	MPP2	99	99	89	90	57	0
05478	MPP1	100	100	93	89	50	0
05479	MnBP	99	100	81	75	0	0
05480	TR	99	98	76	79	48	0
05481	GH36	99	99	78	82	63	0
05482	Pgm	99	100	92	94	94	0
05483	GH113	99	99	76	70	0	0
07683	GH26	99	99	64	0	0	0
07684	GH3B	99	99	62	76	76	0
07685	GH3A	99	99	62	76	76	0

Supplementary Table 3. Homologs of *Ri*GH26 (catalytic domain only)[a,b].

Strain	Accession Number	Coverage	Identity	Length[c]	CBM27	CBM23
R. intestinalis XB6B4	CBL14297.1	100%	99%	1308	Y	Y
R. intestinalis M50/1	CBL07467.1	100%	99%	1314	Y	Y
R. faecis	WP_055067491.1	80%	47%	1374	Y	Y
Roseburia sp. CAG:18_43_25	OLA58885.1	77%	48%	1519	Y	Y
Coprococcus eutactus CAG:665	CCZ93652.1	85%	48%	1555	Y	Y
Coprococcus sp. CAG:131	CDB80322.1	85%	47%	1544	Y	Y
Clostridium butyricum	WP_058146428.1	72%	40%	1444	Y	Y
Clostridium disporicum	CUO64427.1	72%	38%	1333	Y	Y
Clostridium sp. Marseille-P2414	WP_066889682.1	73%	37%	1147	Y	Y
Clostridium beijerinckii	WP_026885856.1	69%	39%	1398	Y	Y
Clostridium saudiense	WP_052330667.1	80%	35%	1148	Y	Y
Clostridium puniceum	WP_077847551.1	64%	38%	1016	Y	Y
Propionispora vibrioides	WP_091746906.1	68%	36%	889	Y	Y
Clostridium sp. DSM 8431	WP_090011638.1	66%	36%	1692	Y	Y
Selenomonas bovis	WP_052177425.1	64%	35%	891	Y	Y
Anaerocolumna aminovalerica	SFO28893.1	64%	35%	928	Y	Y

[a]Homologs were identified by BLAST searches against the NCBI non-redundant protein sequence database. [b]Y indicates the presence of a specific CBM in the protein sequence. [C]Amino Acids

Supplementary Table 4. Homologs of *Ri*CBM27[a].

Strain	Phylum	Accession Number	E- value	Coverage	Identity
R. intestinalis XB6B4	Firmicutes	CBL14297.1	2.00E-101	100%	99%
R. intestinalis M50/1	Firmicutes	CBL07467.1	2.00E-100	100%	99%
Coprococcus eutactus CAG:665	Firmicutes	CCZ93652.1	2.00E-23	100%	41%
Coprococcus sp. CAG:131	Firmicutes	CDB80322.1	8.00E-23	100%	40%
R. faecis	Firmicutes	WP_055067491.1	2.00E-22	90%	41%
Roseburia sp. CAG:18_43_25	Firmicutes	OLA58885.1	3.00E-22	90%	41%
Clostridium disporicum	Firmicutes	CUO64427.1	1.00E-18	98%	34%
Clostridium butyricum	Firmicutes	WP_058146428.1	7.00E-18	93%	39%
Clostridium sp. Marseille-P2414	Firmicutes	WP_066889682.1	2.00E-17	98%	33%
Clostridium puniceum	Firmicutes	WP_077847551.1	4.00E-17	91%	34%
Uncultured *Clostridium sp.*	Firmicutes	SCJ87681.1	3.00E-15	91%	32%
Clostridium saudiense	Firmicutes	WP_052330667.1	4.00E-15	91%	32%
Clostridium beijerinckii	Firmicutes	WP_026885856.1	3.00E-11	98%	33%
Anaerocolumna aminovalerica	Firmicutes	SFO28893.1	6.00E-11	96%	32%
Selenomonas bovis	Firmicutes	WP_052177425.1	9.00E-11	90%	31%
Clostridium sp. DSM 8431	Firmicutes	WP_090011638.1	2.00E-10	98%	30%
Propionispora vibrioides	Firmicutes	WP_091746906.1	8.00E-10	91%	32%
Pelosinus fermentans JBW45	Firmicutes	AJQ29746.1	3.00E-09	92%	31%

[a]Homologs were identified by BLAST searches against the NCBI non-redundant protein sequence database.

Supplementary Table 5. Homologs of *Ri*CBM23[a].

Strain	Phylum	Accession Number	E- value	Coverage	Identity
R. intestinalis XB6B4	Firmicutes	CBL14297.1	1.00E-123	100%	99%
R. intestinalis M50/1	Firmicutes	CBL07467.1	1.00E-123	100%	99%
Coprococcus eutactus CAG:665	Firmicutes	CCZ93652.1	5.00E-58	93%	58%
Coprococcus sp. CAG:131	Firmicutes	CDB80322.1	2.00E-57	93%	58%
R. faecis	Firmicutes	WP_055067491.1	3.00E-44	88%	49%
Roseburia sp. CAG:18_43_25	Firmicutes	OLA58885.1	4.00E-43	88%	49%
Clostridiales bacterium CHKCI001	Firmicutes	CVI73099.1	4.00E-41	93%	44%
Clostridium beijerinckii	Firmicutes	WP_026885856.1	9.00E-40	89%	47%
Clostridium butyricum	Firmicutes	WP_058146428.1	5.00E-36	94%	42%
Propionispora vibrioides	Firmicutes	WP_091746906.1	2.00E-34	99%	43%
Clostridium puniceum	Firmicutes	WP_077847551.1	2.00E-35	89%	41%
Selenomonas bovis	Firmicutes	WP_052177425.1	2.00E-33	89%	42%
Clostridium sp. DSM 8431	Firmicutes	WP_090011638.1	4.00E-33	93%	40%
Clostridium disporicum	Firmicutes	WP_055276275.1	2.00E-30	94%	41%
Clostridium sp. Marseille-P2414	Firmicutes	WP_066889682.1	3.00E-30	91%	40%
Clostridium saudiense	Firmicutes	WP_052330667.1	5.00E-30	96%	41%
Clostridium populeti	Firmicutes	WP_092560634.1	2.00E-21	92%	33%
Lachnospiraceae bacterium ND2006	Firmicutes	WP_051666040.1	2.00E-21	100%	34%
Cohnella sp. OV330	Firmicutes	WP_090116078.1	1.00E-18	89%	36%
Anaerocolumna aminovalerica	Firmicutes	SFO28893.1	3.00E-17	99%	32%

[a]Homologs were identified by BLAST searches against the NCBI non-redundant protein sequence database.

Supplementary Table 6. Carbohydrates used in this study, listed with supplier and product number.

Substrate	Source	Catalogue Number
Glucose	Sigma	(G8270)
Galactose	Sigma	(G0750)
Mannose	Sigma	(M4625)
Mannobiose	Megazyme	(O-MBI)
Mannotriose	Megazyme	(O-MTR)
Mannotetraose	Megazyme	(O-MTE)
Mannopentaose	Megazyme	(O-MPE)
Mannohexaose	Megazyme	(O-MHE)
Cellobiose	Sigma	(C7252)
Cellotriose	Megazyme	(O-CTR)
Cellotetraose	Megazyme	(O-CTE)
Cellopentaose	Megazyme	(O-CPE)
Cellohexaose	Megazyme	(O-CHE)
Mannose-1-phosphate	Sigma	(M1755)
Mannose-6-phosphate	Sigma	(M6876)
Glucose-1-phosphate	Sigma	(G7000)
Glucose-6-phosphate	Sigma	(G7879)
Fructose-6-phosphate	Sigma	(F3627)
Glucosylmannose plus Mannobiose	Megazyme	(O-GMMBI)
6^1-α-D-Galactosylmannobiose plus Mannotriose	Megazyme	(OGMM-3)
6^1-α-D-Galactosylmannotriose	Megazyme	(O-GM3)
6^3,6^4-α-D-Galactosylmannopentaose	Megazyme	(O-GGM5)
Galactomannan (guar gum)	Sigma	(G4129)
Galactomannan (carob)	Megazyme	(P-GALML)
Glucomannan (konjac)	Megazyme	(P-GLCML)
Acetylated galactoglucomannan (Spruce)	This	
Birch xylan	This	
Lichenan (icelandic moss)	Megazyme	(P-LICHN)
β-glucan (barley)	Megazyme	(P-BGBL)
Curdlan	Megazyme	(P-CURDL)
Wheat arabinoxylan	Megazyme	(P-WAXYL)

#Produced at the Bioprocess Technology and Biorefining Lab, Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, according to procedure described in[1].

Supplementary Table 7. Primers, vector for overexpression in *E. coli* and protein purification techniques used in this study[a].

Gene	Primer (5' - 3')	Vector	Purif. Tech.[b]
ROSINTL182_07683 (CBM27/GH26/CBM23)	**F:**TTAAGAAGGAGATATACTATGGGGGGATACGAGTATGTCTATGC **R:**AATGGTGGTGATGATGGTGCGCTCCACCTTCTGTCACATTTAAC	pNIC-CH	IMAC/SEC
ROSINTL182_07683 (CBM27)	**F:**TTAAGAAGGAGATATACTATGGGGGGATACGAGTATGTCTATGC **R:**AATGGTGGTGATGATGGTGCGCGCCCGCTGCATTATACTGG	pNIC-CH	IMAC/SEC
ROSINTL182_07683 (CBM23)	**F:**TTAAGAAGGAGATATACTATGTCTGTTTTTGCGAATGGAACAAAC **R:**AATGGTGGTGATGATGGTGCGCTCCGGACACGAGTGACTGATC	pNIC-CH	IMAC/SEC
ROSINTL182_07684	**F:**TTAAGAAGGAGATATACTATGGGTAACAGTGGAATCGGTGTTC **R:**AATGGTGGTGATGATGGTGCGCCTCTAACTCAATCCAGTCCAACTTC	pNIC-CH	IMAC/SEC
ROSINTL182_07685	**F:**TTAAGAAGGAGATATACTATGGAAAAATGGCAAAGATCACTTTATC **R:**AATGGTGGTGATGATGGTGCGCGTCGATTTTCAGTATTAATCCTAAAATAC	pNIC-CH	IMAC/SEC
ROSINTL182_05469 (GH1_D2)	**F:**TTAAGAAGGAGATATACTATGATACTTAGTATGAATCAGGCAAATAAAG **R:**AATGGTGGTGATGATGGTGCGCTTTTACAGTAGAAACGATCAGACTCATC	pNIC-CH	IMAC/SEC
ROSINTL182_05470 (GH1_D1)	**F:** TTAAGAAGGAGATATACTATGTTAGACCCGGTATGTACGCAC **R:** AATGGTGGTGATGATGGTGCGCTCCTCCATTCGTTTCTATTACTTTCTG	pNIC-CH	IMAC/SEC
ROSINTL182_05471	**F:**TTAAGAAGGAGATATACTATGGAATATCAAATTAAATACGAAAACGGC **R:**AATGGTGGTGATGATGGTGCGCTTTTTCAGAGGAACCAATGACAGAC	pNIC-CH	IMAC/SEC
ROSINTL182_05473	**F:**TTAAGAAGGAGATATACTATGAAACGTGTGATGGAGTGTCCG **R:**AATGGTGGTGATGATGGTGCGCAGATTCCCAGACTGCATCCC	pNIC-CH	IMAC/SEC
ROSINTL182_05474	**F:**TTAAGAAGGAGATATACTATGAGCAAATATAAAATGATCAGCCAGC **R:**AATGGTGGTGATGATGGTGCGCGCGTCTTCCGATCTCGGTATC	pNIC-CH	IMAC/SEC
ROSINTL182_05475	**F:**TTAAGAAGGAGATATACTATGGAGATTCGACGAATGTTACATGAG **R:**AATGGTGGTGATGATGGTGCGCTTTATCTTCCTTTGCAAGGAACTC	pNIC-CH	IMAC/SEC
ROSINTL182_05476	**F:**TTAAGAAGGAGATATACTATGAGTGAGTTAAAAGAACTTGCAGCAG **R:**AATGGTGGTGATGATGGTGCGCGAATCAGTTTTATGCTCCGATTCTG	pNIC-CH	HIC/ SEC
ROSINTL182_05479	**F:** TTTCAGGGCGCCATGGGCTCAAACAACACAGCAGG **R:** GACGGAGCTCGAATTTTACTCAGTTACTTCAATTCCGTTTG	pETM-11	IMAC/SEC
ROSINTL182_05481	**F:**TTAAGAAGGAGATATACTATGGCAATCACATACTTTGAAAAAGAACGC **R:**AATGGTGGTGATGATGGTGCGCTTTTTCCATTTAATCTTATTTCTTTTCCAC	pNIC-CH	HIC/ SEC
ROSINTL182_05482	**F:**TTAAGAAGGAGATATACTATGAGCAATTACATGGAAACATACAAACAG **R:**AATGGTGGTGATGATGGTGCGCTTCCTACTGTTATTTCAGCAAGTCTTC	pNIC-CH	HIC/ SEC
ROSINTL182_05483	**F:**TAAGAAGGAGATATACTATGGAATATATCTGTGGGGTGACTTTTGC **R:**AATGGTGGTGATGATGGTGCGCTTGACCTCAACATCATTTATTAAAAATATC	pNIC-CH	HIC/ SEC

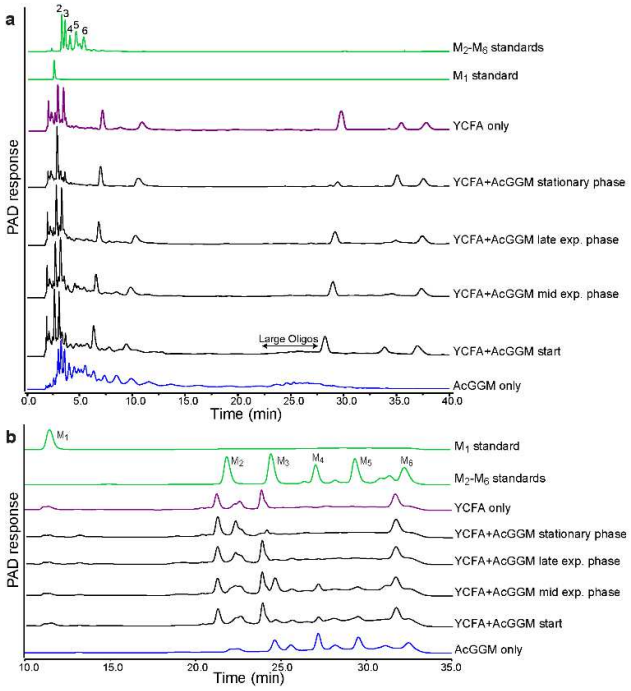
[a]The 5' extension sequences used for molecular cloning are underlined. [b]Purification techniques employed.

Supplementary Table 8. Primers used for sequencing of the intergenic region between ROSINTL182_05470 - 05469.

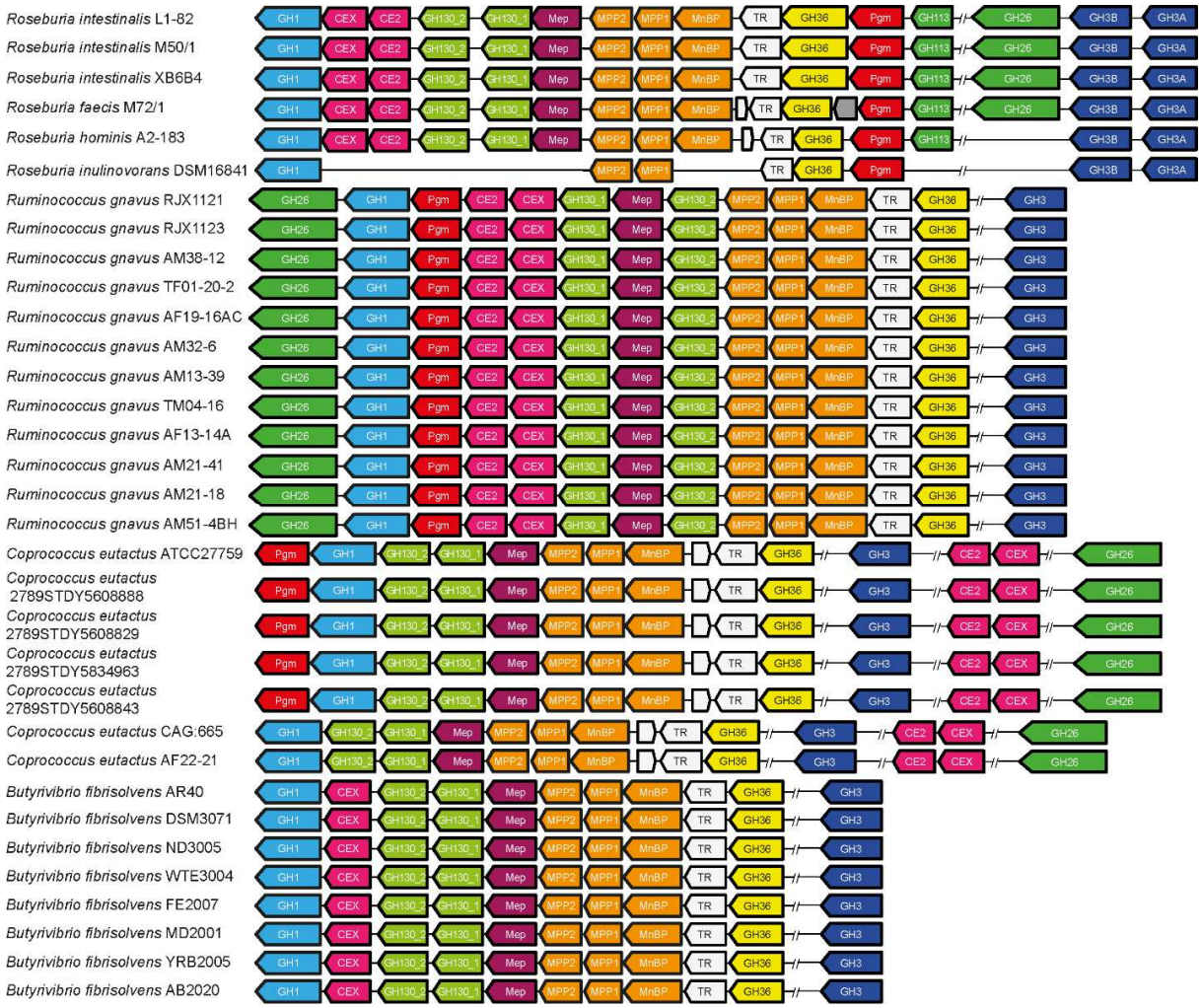
Gene	Primer (5' - 3')	Application
ROSINTL182_05470/ ROSINTL182_05469	**F:**GGGATTACGCCATATATTACC **R:**GTCTCCGTTTGGATGTGC	Sequencing of the intergenic region between *Ri*GH1_D1 and *Ri*GH1_D2

Supplementary Table 9. Primers used for qPCR.

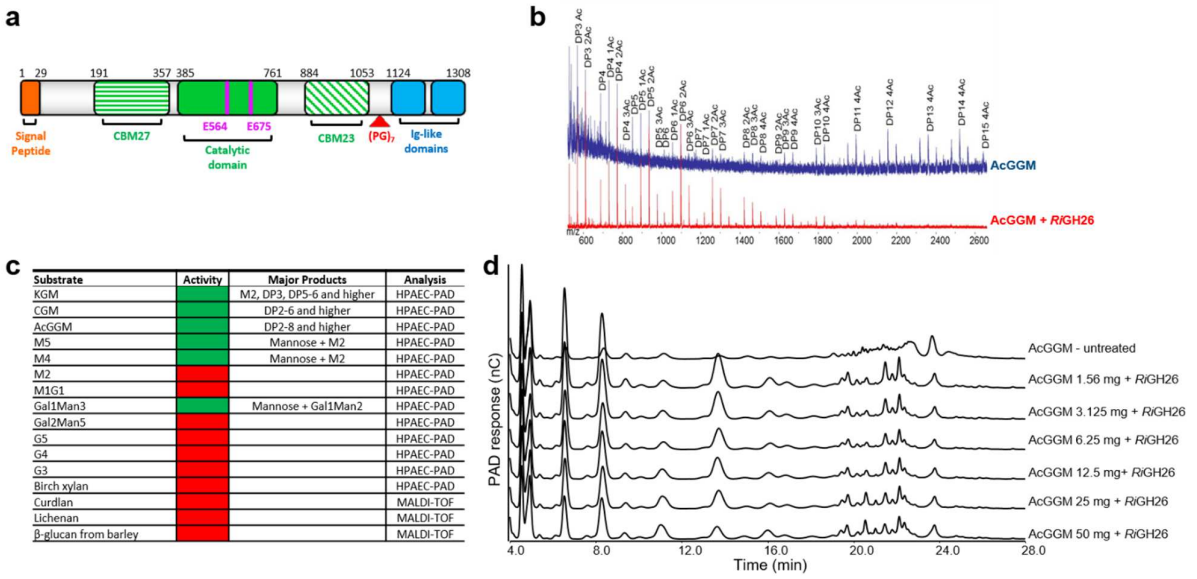
Target	Primer (5' - 3')	Supplementary Reference
R. intestinalis L1-82	**F:** GCATGACCTGGTGTGAA **R:** TTGGGCCGTGTCTCA	Larsen, N. et al[2]
B. ovatus ATCC 8384	**F:** GGTGTCGGCTTAAGTGCCAT **R:** CGGAYGTAAGGGCCGTGC	Larsen, N. et al[2]
Universal 16S rRNA	**F:** CCTAYGGGRBGCASCAG **R:** GGACTACNNGGGTATCTAAT	Yu, Y. et al[3]



Supplementary Figure 1. *R. intestinalis* L1-82 AcGGM consumption at zero, middle and later time points, determined by HPAEC-PAD. a, Chromatograms showing the oligosaccharides detected in the supernatant of *R. intestinalis* L1-82 while growing on YCFA supplemented with 5 mg ml^{-1} AcGGM. M_2-M_6 standards were 2, mannobiose; 3, mannotriose; 4, mannotetraose; 5, mannopentaose; 6, mannohexaose. **b**, Chromatograms with a gradient adapted to improve the resolution of mono-, di- and shorter oligosaccharides present in the same supernatant as in a. The data are representative of three biological replicates. Abbreviations: mid exp. phase; middle exponential phase; late exp. phase, late exponential phase.

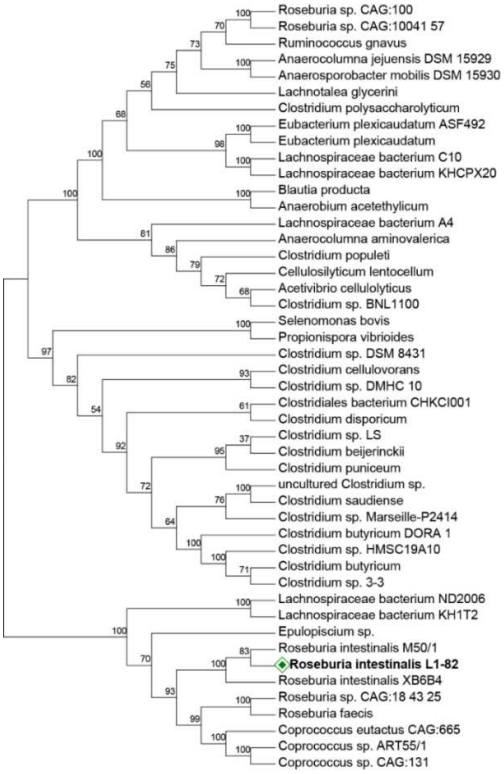


Supplementary Figure 2. Sequence conservation and genomic organization of the β-mannan utilization loci among *Roseburia* species and closely related members of the Clostridium cluster XIVa. Common predicted functions are colored according to Fig. 2d. The figure shows that MULL and MULS are conserved in all the *R. intestinalis* strains, *R. faecis* and *R. hominis* A2-183; however, the latter genome does not include an orthologous GH26 gene. *R. inulinovorans* lacks most of the genes encoding the β-mannan utilization apparatus, suggesting a different carbohydrate utilization specialization for this strain. Among members of the Clostridium cluster XIVa, several *R. gnavus* and *C. eutactus* strains harbor similar MULL and MULS genes, although no GH113 was detected. The mannanase GH26 is lacking from the genomes of the *B. fibrisolvens* strains, suggesting that these bacteria are likely unable to hydrolyze polymeric β-mannans.

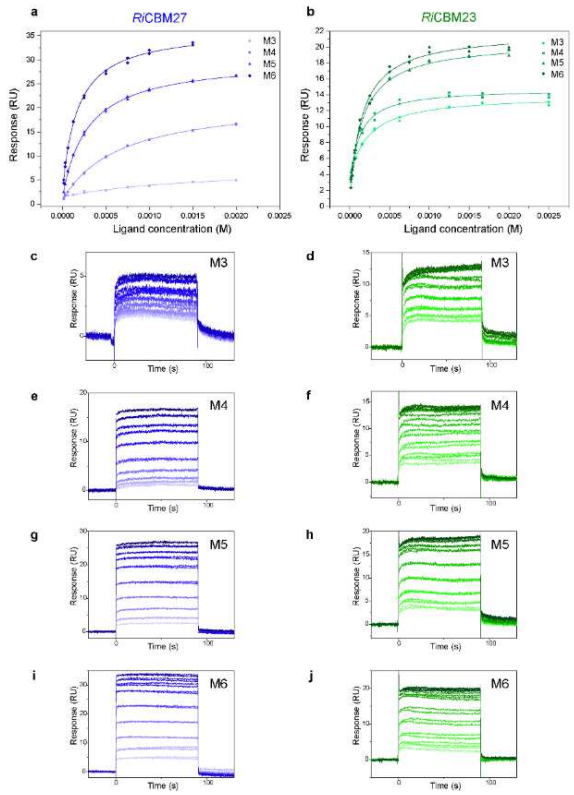


Substrate	Activity	Major Products	Analysis
KGM		M2, DP3, DP5-6 and higher	HPAEC-PAD
CGM		DP2-6 and higher	HPAEC-PAD
AcGGM		DP2-8 and higher	HPAEC-PAD
M5		Mannose + M2	HPAEC-PAD
M4		Mannose + M2	HPAEC-PAD
M2			HPAEC-PAD
M1G1			HPAEC-PAD
Gal1Man3		Mannose + Gal1Man2	HPAEC-PAD
Gal2Man5			HPAEC-PAD
G5			HPAEC-PAD
G4			HPAEC-PAD
G3			HPAEC-PAD
Birch xylan			HPAEC-PAD
Curdlan			MALDI-TOF
Lichenan			MALDI-TOF
β-glucan from barley			MALDI-TOF

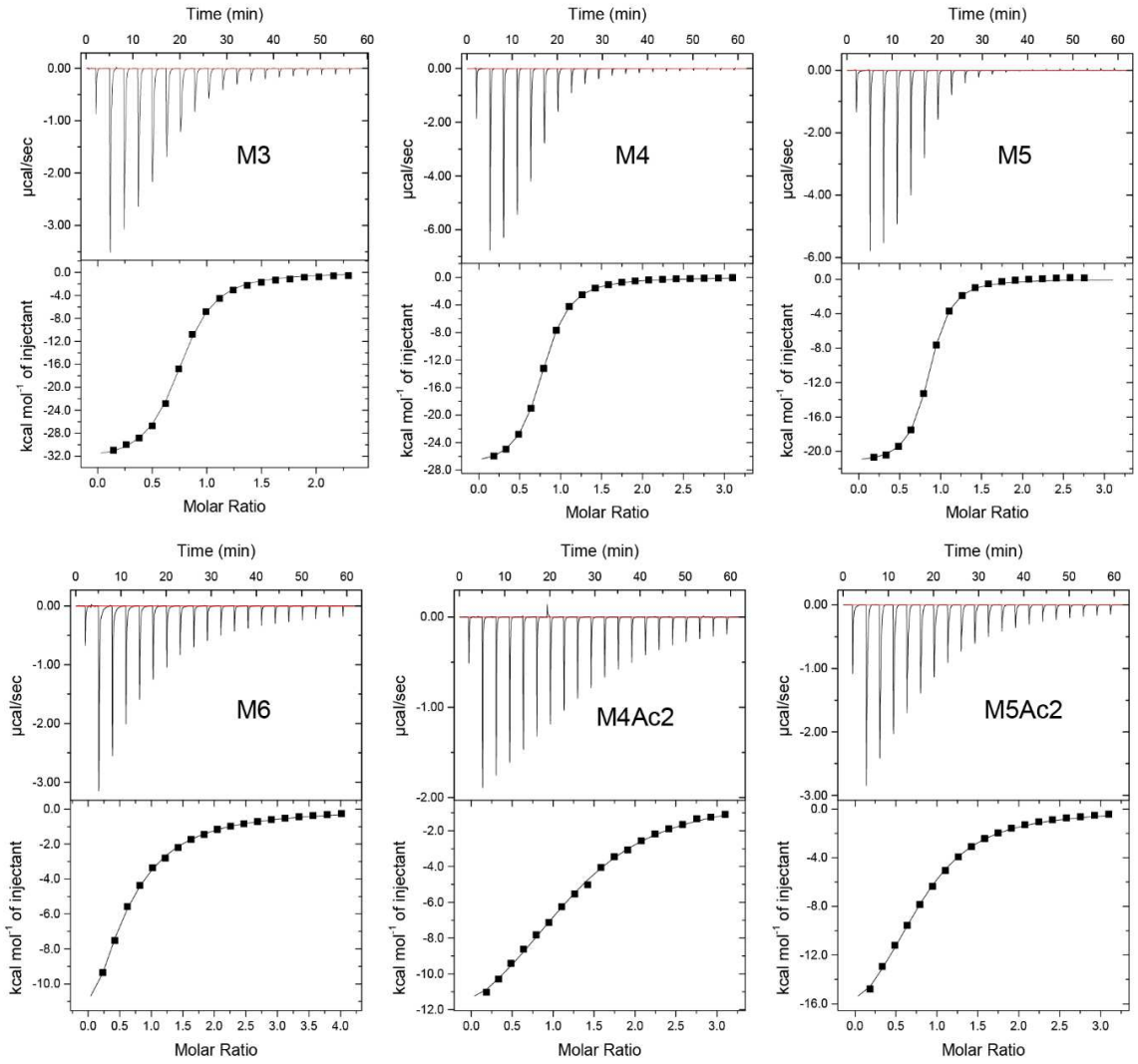
Supplementary Figure 3. Schematic of *Ri*GH26 and activity on different glycans. a, Domain organization of *Ri*GH26. The protein contains an N-terminal secretion signal peptide (in orange), as detected by the SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP/), which is predicted to be removed by cleavage between Gly29 and Tyr30. The CBMs and catalytic domain are shown in green with the catalytic residues in magenta. The red arrow indicates the position of the Proline-Glycine (PG) rich region. The two Ig-like domains are shown in light blue. Based on these predictions, the gene was cloned and the protein expressed without the N-terminal signal, C-terminal PG-rich region and Ig-like sequences. **b**, Activity of *Ri*GH26 on AcGGM. The picture shows the MALDI-TOF spectra of oligosaccharides before (blue) and after *Ri*GH26 treatment (red). **c**, Summary of the activity of *Ri*GH26 on different glycans under standard assay conditions. A green box indicates activity while a red box indicates the absence of detectable products. **d**, Activity of *Ri*GH26 on different concentrations of AcGGM. *Ri*GH26, at 10 nM, was incubated with increasing concentration of AcGGM for 1 h in standard assay conditions. The reaction products were diluted to have a final concentration of approximately 1 mg ml^{-1} and subjected to HPAEC-PAD.



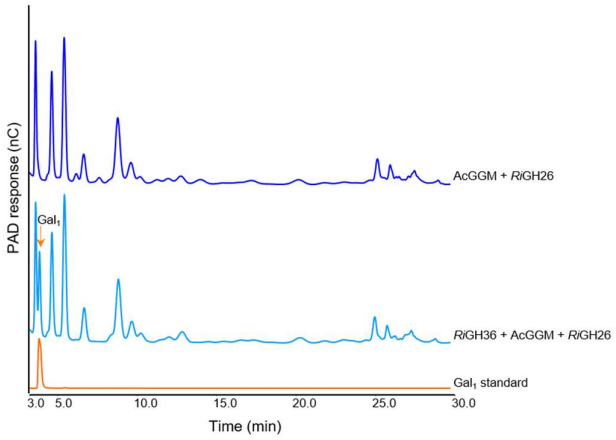
Supplementary Figure 4. Unrooted maximum likelihood tree of *Ri*GH26 and homologs. Searching for *Ri*GH26-like proteins was performed by BLAST against the NCBI non-redundant protein sequence database. Sequences with coverage <86% and identity <40% were eliminated. Multiple sequence alignment was performed using ClustalW. The phylogenetic tree was generated using the maximum likelihood algorithm in Mega 7 (http://megasoftware.net/). Bootstrap values are shown at each node as percentage of 500 replicates. *Ri*GH26 is indicated by a green-colored diamond. All sequences were identified as members of the Firmicutes phylum, with 96% of the sequence belonging to the order Clostridiales. Exceptions were the sequence of *Selenomonas bovis* and *Propionospora vibrioides* that belong to the class Negativicutes. The majority of the sequences were affiliated to the Clostridium Cluster XIVa.



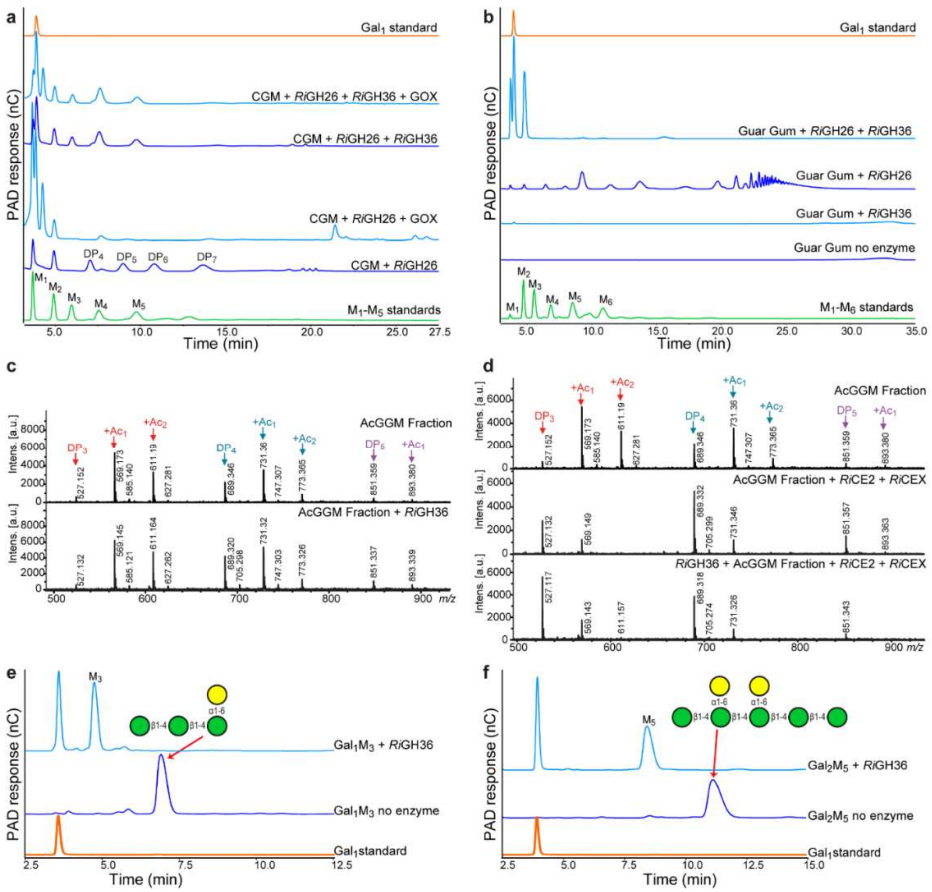
Supplementary Figure 5. Binding of *Ri*CBM27 and *Ri*CBM23 to manno-oligosaccharides using SPR analysis (related to Table 1). Blank and reference cell corrected (**a**-**b**) and raw SPR sensograms (**c**-**j**) showing the binding of *Ri*CBM27 (blue) and *Ri*CBM23 (green) to manno-oligosaccharides (M3-6). In all panels, the results are representative of at least two independent replicates.



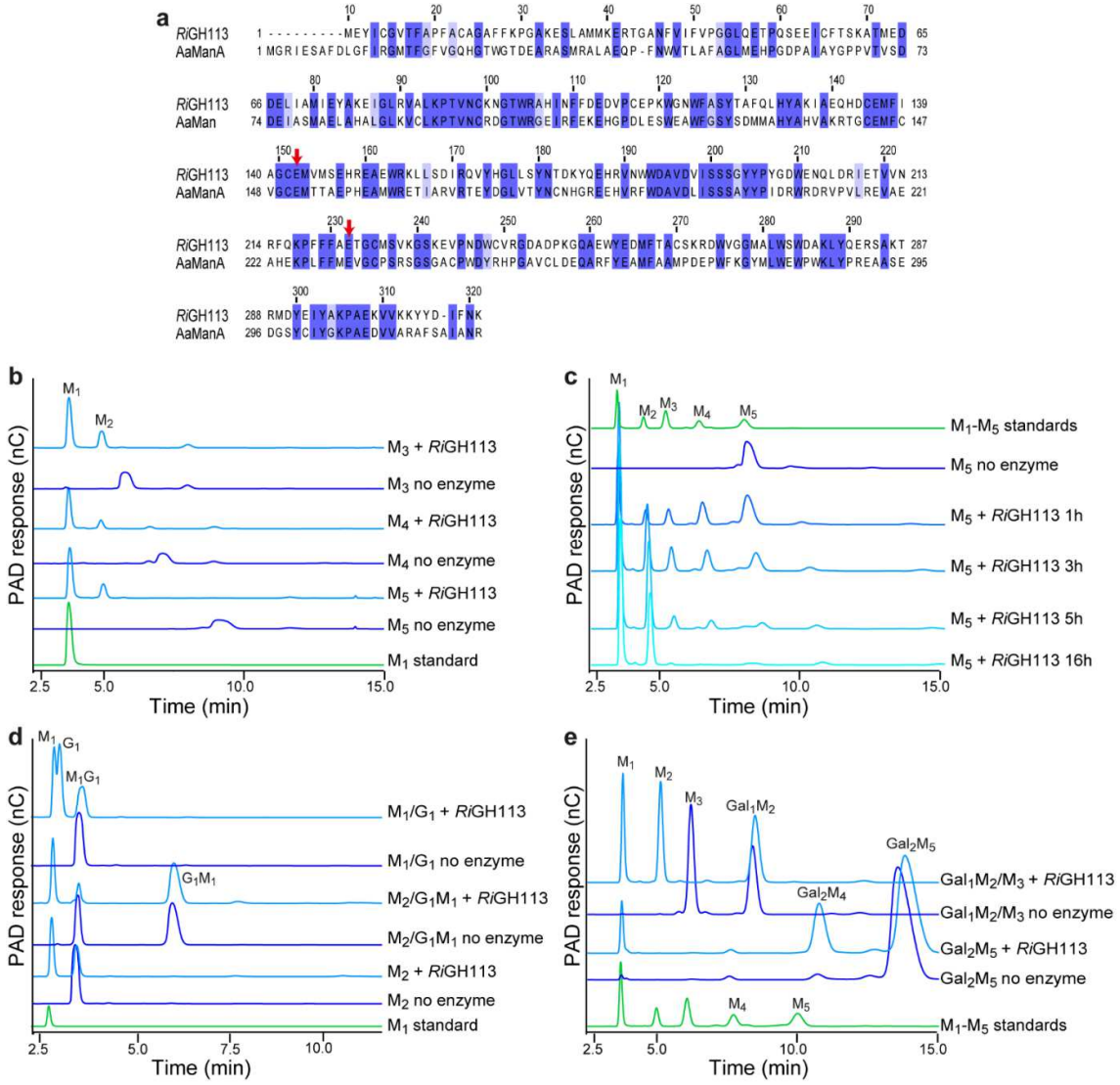
Supplementary Figure 6. ITC analysis of *Ri*MnBP binding to linear and substituted manno-oligosaccharides (related to Table 2). Data show *Ri*MnBP binding to manno-oligosaccharides (M3-6), diacetylated mannotetraose (M4Ac2) and diacetylated mannopentaose (M5Ac2). The top panel in each pair shows the thermograms, whereas the bottom graph depicts the binding isotherms and one set of equivalent binding sites model fits to the data (solid lines).



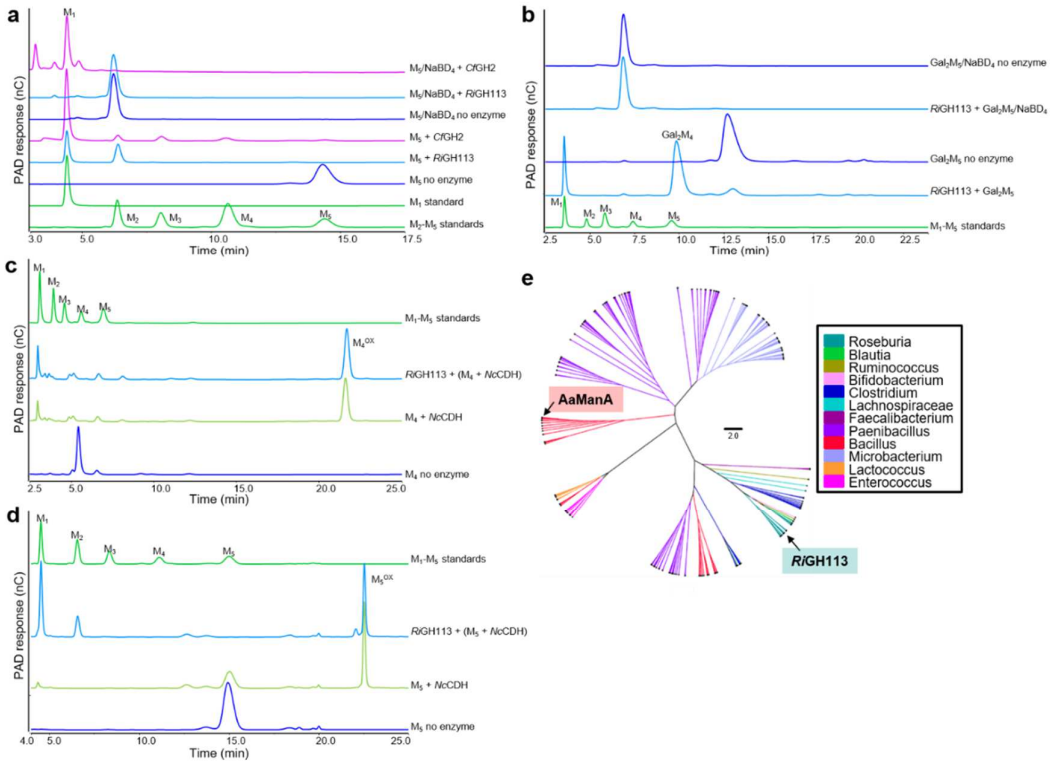
Supplementary Figure 7. Activity of *Ri*GH36 on AcGGM. HPAEC-PAD analysis of *Ri*GH26-treated AcGGM before and after subsequent incubation with *Ri*GH36.



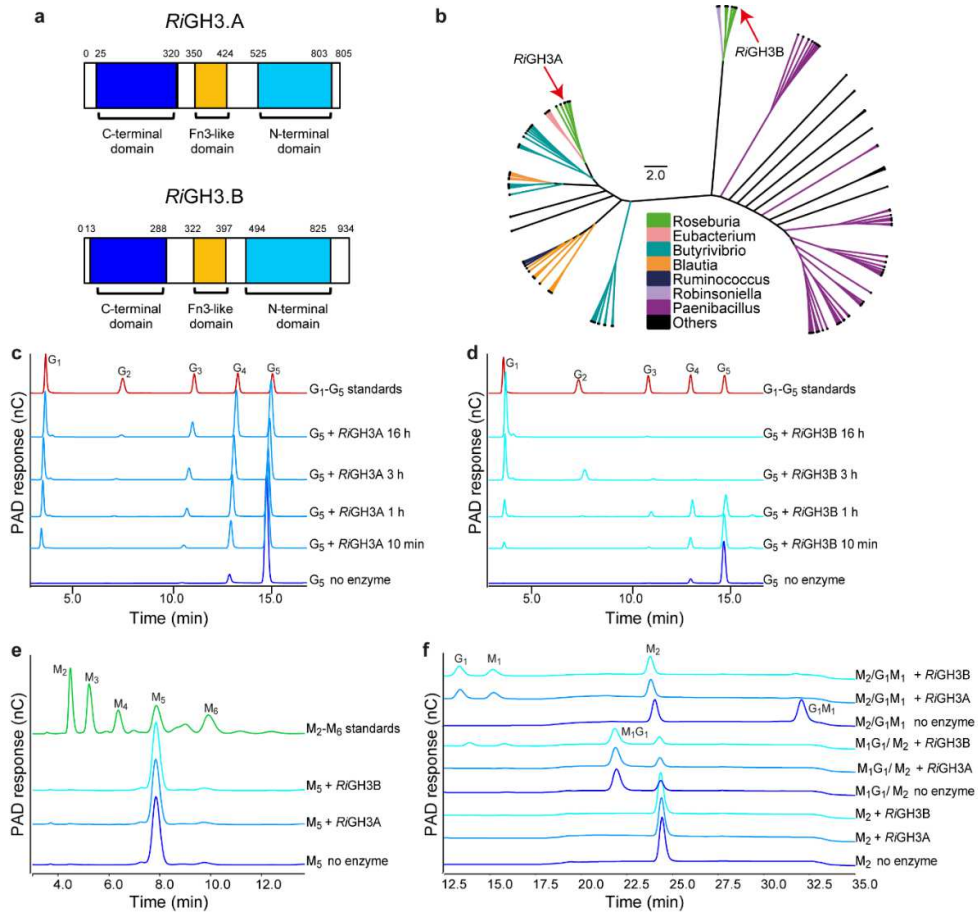
Supplementary Figure 8. Activity of *Ri*GH36 on galactomannan, acetylated manno-oligosaccharides and galactomanno-oligosaccharides. a, HPAEC-PAD chromatograms showing the effect of the treatment with galactose oxidase (GOX from *Dactylium dendroides*, Megazyme) on CGM pre-hydrolyzed with *Ri*GH26 and subsequent incubation with *Ri*GH36. The oligosaccharides (DP4-7) released from CGM by *Ri*GH26 were oxidized by the GOX; indeed, upon GOX treatment, the galactosylated oligosaccharides at denoted DP4-7 disappear and the corresponding oxidized oligosaccharides appear at 20+ minutes. No oxidized product could be observed after GOX-treatment of the *Ri*GH26 and *Ri*GH36-treated CGM sample, demonstrating that *Ri*GH36 completely removes the galactose residues from the substrates. **b**, HPAEC-PAD analysis of the products released from guar gum galactomannan by *Ri*GH36 alone or when the enzyme is combined with the β-mannanase *Ri*GH26. The figure shows that *Ri*GH36 has limited activity on the polymeric substrate while it acts synergistically with *Ri*GH26 to degrade the guar gum galactomannan completely into mannose, galactose and mannobiose. **c**, MALDI-TOF spectra of the hydrolysis of a preparation containing trisaccharides (DP3), tetrasaccharides (DP4) and pentasaccharides (DP5) with different degree of acetylation (Ac), after overnight incubation with *Ri*GH36. **d**, Increased *Ri*GH36 activity can be observed as an increase in DP3 and a corresponding reduction in DP4 and DP5 when the same substrate as in panel c was pretreated with *Ri*CE2 and *Ri*CEX to remove the acetyl substitutions. **e**, HPAEC-PAD analysis of the reaction products generated after *Ri*GH36 treatment of Gal_1M_3 (15 mM). **f**, Products released from Gal_2M_5 (8.7 mM) by *Ri*GH36 action as analyzed by HPAEC-PAD. In all panels, *Ri*GH36 was used at a concentration of 15 nM and incubated with the glycans for 16 h at 37 °C.



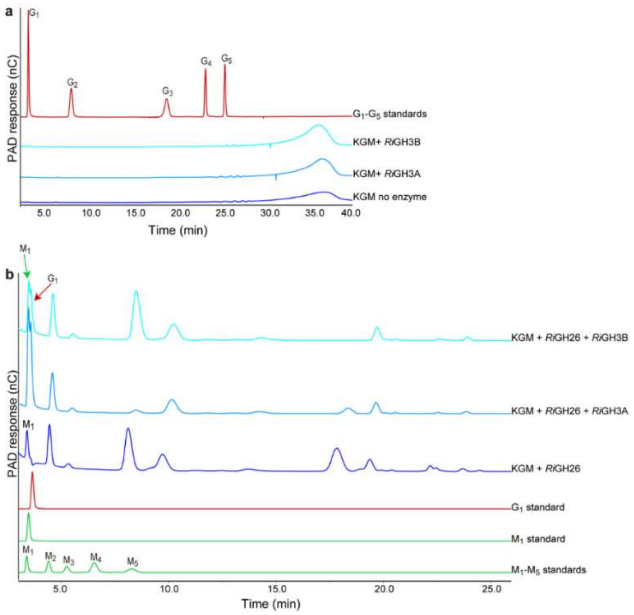
Supplementary Figure 9. Sequence analysis, activity of *Ri*GH113 towards various manno-oligosaccharides and time course activity. a, Amino acid sequence alignment of the GH113 protein of *R. intestinalis* L1-82 (*Ri*GH113) and AaManA from *Alicyclobacillus acidocaldarius*. The protein sequences were aligned with JalView (http://www.jalview.org) and conserved amino acids are shown in blue. Red arrows indicate the catalytic and substrate interacting residues. **b**, Pattern of hydrolysis of manno-oligosaccharides by *Ri*GH113. Enzyme assays were performed overnight. **c**, Pattern of hydrolysis of M_5 by *Ri*GH113. Samples were withdrawn at the indicated time, enzymes were inactivated using 100 mM NaOH and the hydrolysates were analyzed by HPAEC-PAD. **d**, Reaction products of disaccharides catalyzed by *Ri*GH113. **e**, Hydrolysis of Gal_1M_2 plus M_3 and Gal_2M_5 by *Ri*GH113. The reaction products were analyzed by HPAEC-PAD. Gal_1M_4, indicates galactosylmannotetraose. In all panels, control reactions without enzyme did not yield products.



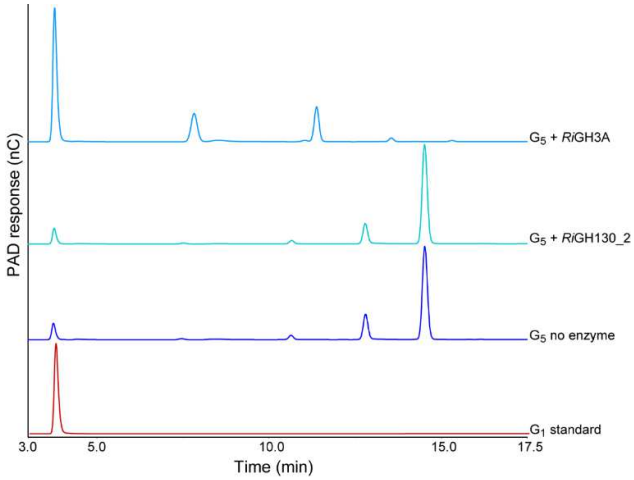
Supplementary Figure 10. Phylogeny and inhibition of *Ri*GH113 activity when blocking the reducing end of various manno-oligosaccharides. a, Analysis of *Ri*GH113 activity on native and reduced M_5. A commercial *Cellulomonas fimi* GH2 β-mannosidase (*Cf*GH2, Megazyme), which acts from the non-reducing end sugar, was used as control. M_5 was pretreated with $NaBD_4$ to reduce the reducing (downstream) end mannose unit into mannitol. *Cf*GH2 fully hydrolyzes the reduced mannopentamer, whereas *Ri*GH113 is not able to hydrolyze it. Note that the reduced oligosaccharide has a considerably shorter retention time than its corresponding native oligosaccharide. **b**, *Ri*GH113 hydrolysis of native and oxidized M_4 and **c**, M_5 analyzed with HPAEC-PAD. The substrates were pretreated with a *Neurospora crassa* cellobiose dehydrogenase (*Nc*CDH), which oxidizes the reducing end monosaccharide unit into the corresponding lactone/aldonic acid, resulting in a considerably longer retention time than its corresponding native oligosaccharide. *Ri*GH113 is not able to hydrolyze the oxidized oligosaccharide. **d**, HPAEC-PAD analysis of the activity of *Ri*GH113 on native and reduced Gal_2Man_5. Gal_2Man_5 was treated with $NaBD_4$ before addition of *Ri*GH113. **e**, Phylogenetic tree of *Ri*GH113 and homologs identified by a Blast search against the NCBI non-redundant protein database. Sequences with coverage ≥86% and identity ≥40% were selected. The resulting 298 sequences were aligned using Muscle. A phylogenetic tree was generated using the maximum likelihood algorithm with 500 bootstrap repetitions in Mega 7 (http://megasoftware.net/). The tree was visualized using Figtree (http://tree.bio.ed.ac.uk/software/figtree). *Ri*GH113 and AaManA are indicated with a black arrow. All sequences were identified as members of the Firmicutes phylum with the exception of those belonging to Microbacterium (Actinobacteria).



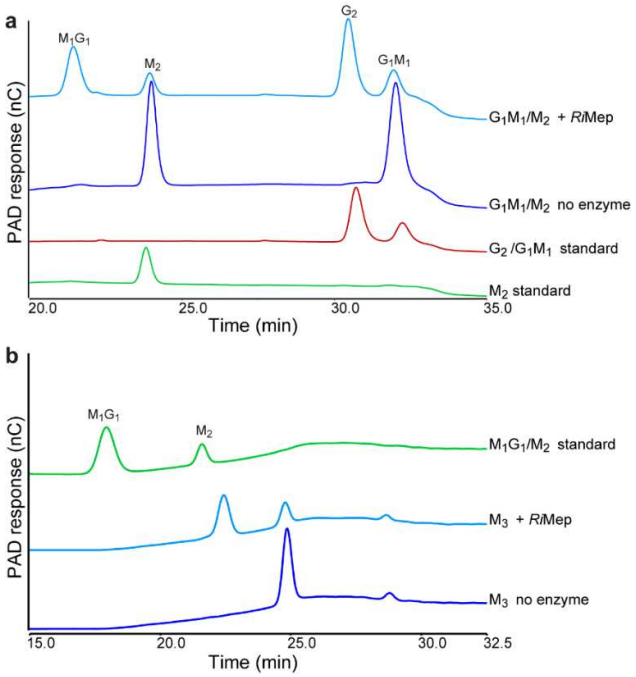
Supplementary Figure 11. Phylogeny of *Ri*GH3A, *Ri*GH3B and homologs and β-glucosidases degradation of different oligosaccharides over time. a, Domain organization of *Ri*GH3A and *Ri*GH3B. Similar to other β-glucosidases from Clostridiales[4], the N- and C-terminal modules in both *Ri*GH3A and *Ri*GH3B are arranged in reverse sequence. **b**, Phylogenetic tree of *Ri*GH3A, *Ri*GH3B and homologs identified by BlastP search against the NCBI non-redundant protein database. Sequences with coverage <86% and identity <40% were removed. The resulting 183 sequences were aligned using Muscle. A phylogenetic tree was generated using the maximum likelihood algorithm with 500 bootstrap repetitions in Mega 7 (http://megasoftware.net/). The tree was visualized using Figtree (http://tree.bio.ed.ac.uk/software/figtree). *Ri*GH3A and *Ri*GH3B are indicated with a red arrow. Despite an apparent redundancy in structure and biochemical function, the two β-glucosidases appear to have diverged significantly, sharing only 40% identity at the amino acid level. Time course analysis of enzymatic reactions containing **c**, *Ri*GH3A and **d**, *Ri*GH3B. Equal amount of enzymes (10 nM) were used. Aliquots were taken at the indicated time points and the enzymes were inactivated by adding NaOH to 100 mM. **e**, HPAEC-PAD analysis of M_5 incubated overnight in the absence or presence of *Ri*GH3A and *Ri*GH3B, showing that neither of them are active on this substrate. **f**, Activity of the β-glucosidases *Ri*GH3A and *Ri*GH3B on disaccharides.



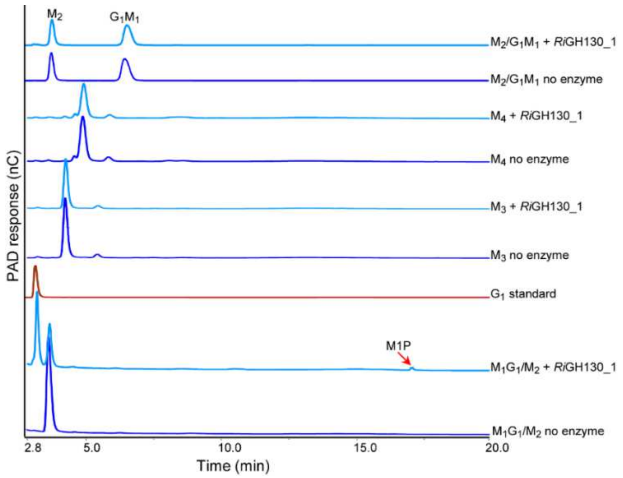
Supplementary Figure 12. *Ri*GH3A and *Ri*GH3B activity on glucomannan. **a**, HPAEC-PAD analysis of *Ri*GH3A or *Ri*GH3B activity on KGM. **b**, Reaction products of *Ri*GH26-digested KGM catalyzed by *Ri*GH3A or *Ri*GH3B.



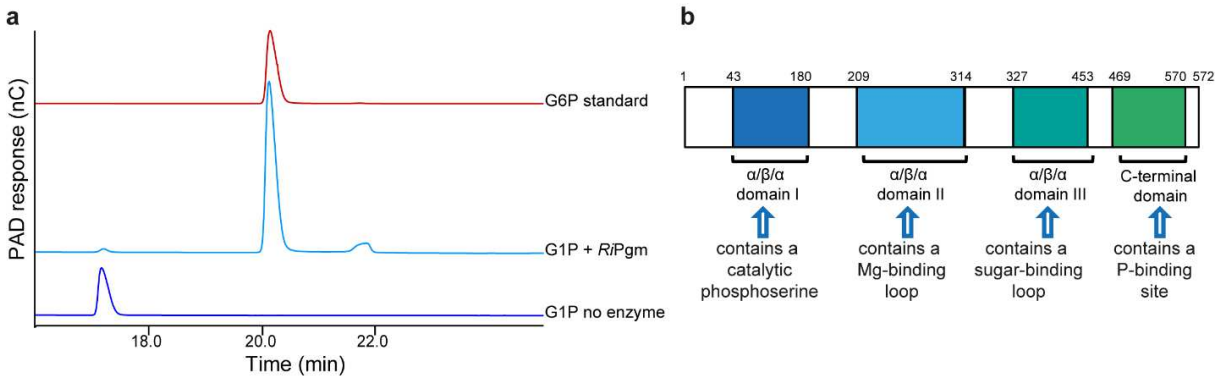
Supplementary Figure 13. Control reaction to check for the activity of *Ri*GH130_2 on cello-oligosaccharides. HPAEC-PAD analysis of G_5 incubated at standard assay conditions in the absence or presence of *Ri*GH130_2. A reaction with the β-glucosidase *Ri*GH3A was used as a control. Only the incubation of *Ri*GH3A with G_5 released a reaction product.



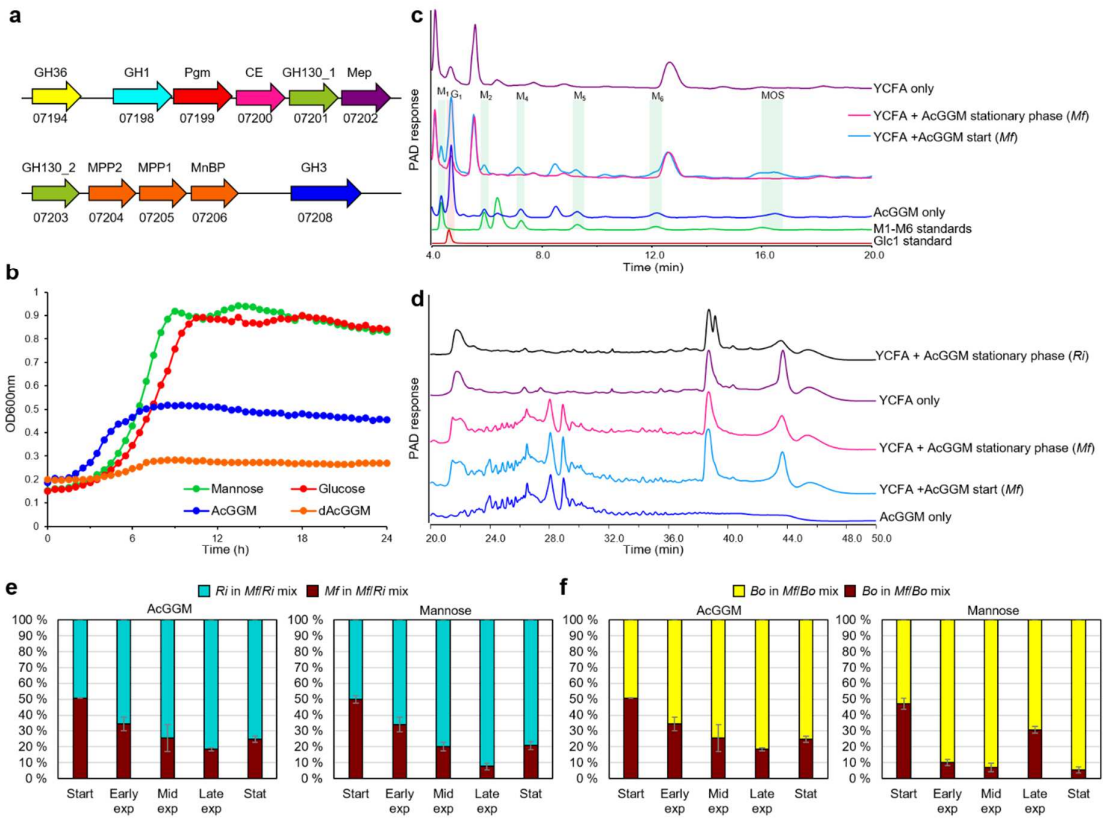
Supplementary Figure 14. HPAEC chromatograms to verify the activity of *Ri*Mep on manno-oligosaccharides. *Ri*Mep was incubated with either **a**, disaccharides or **b**, M_3. *Ri*Mep exhibited epimerization activity towards G_2 and the following manno-oligosaccharides: M_1G_1, G_1M_1, M_3.



Supplementary Figure 15. Control reactions to verify that *Ri*GH130_1 is active only on M_1G_1. HPAEC-PAD analysis of M_4, M_3, a mixture of M_2 and either M_1G_1 or G_1M_1 incubated under standard assay conditions in the absence or presence of *Ri*GH130_1. The reaction products were subjected to HPAEC and identified by their co-migration with known standards. In this elution method, M_1G_1 co-eluted with M_2.



Supplementary Figure 16. Activity on G1P and domain mapping of *Ri*Pgm. a, HPAEC-PAD trace showing the activity of *Ri*Pgm on G1P. G1P was either untreated (G1P no enzyme) or treated with *Ri*Pgm. The G6P produced by *Ri*Pgm was identified by co-migration with the G6P standard. **b**, Domain organization of *Ri*Pgm. As observed for other phosphohexomutases[5], the *Ri*Pgm protein structure is modular, featuring four domains. The catalytic domain spans from residue 43 to 180 and harbors a conserved active site. Downstream to the catalytic core domain are three additional domains: domain II (residues 209-314) contains a conserved magnesium (Mg) -binding loop; domain III (residues 327-453) contains the sugar-binding loop that participates in recognizing the two binding orientations of the 1- and 6-phospho-sugars; and the C-terminal domain IV (residues 469-570) harbors a phosphate (P) -binding site that is required for accommodating the incoming phospho-sugar substrate.



Supplementary Figure 17. *M. formatexigens* β-manno-oligosaccharides utilization cluster, growth profile, glycan consumption and competition experiments. a, Schematic presentation of the putative *M. formatexigens* DSM 14469 β-manno-oligosaccharides degradation locus. Genes are indicated by their locus tag (BRYFOR_XXXXX is abbreviated with the last numbers after the underscore), assignment to CAZyme families and other predicted functions and they are colored according to Fig. 2d. None of the genes has a signal peptide as predicted by the SignalP 4.1 server, suggesting an intracellular location. Absence of a predicted extracellular endomannanase suggests that *M. formatexigens* has no hydrolytic capabilities toward polymeric mannans. **b**, *M. formatexigens* growth on YCFA containing mannose (green), glucose (red), AcGGM (blue) or dyalized AcGGM (dAcGGM, dyalized with a Spectra/Por dialysis tubing 1KDa MW, Repligen, USA) at 0.5% (w/v). The growth experiment was conducted in 200 µl cultures in 96-well microtiter plates incubated at 37 °C in anaerobic conditions. Growth was assessed by measuring the absorbance at 600 nm (OD600 nm) at 15 min intervals using a Powerwave HT absorbance reader (Biotek Instruments, Winooski, VT). HPAEC-PAD traces showing **c**, mono- and oligo-saccharides and **d**, polysaccharides contained in the YCFA supplemented with AcGGM before (turquoise) and after (pink) fermentation with *M. formatexigens* (*Mf*). In panel **c**, MOS indicates manno-oligosaccharides. In panel **d**, samples were chromatographed with the spent supernatant of *R. intestinalis* (*Ri*) after overnight growth on YCFA supplemented with AcGGM at 0.5% (w/v). The data displayed are examples from three biological replicates. The traces show that *M. formatexigens* can utilize oligosaccharides (in panel c), but not the polymeric fraction (in panel d). Relative strain abundance during growth of co-cultures of *M. formatexigens* (*Mf*, dark red) and either **e**, *R. intestinalis* (*Ri*, turquoise) or **f**, *B. ovatus* (*Bo*, yellow) on AcGGM or mannose . The ratio of strains in the co-culture was determined by qPCR using primers described in[6]. Histogram bars show the mean of a biological triplicate. Error bars represent s.d.

SUPPLEMENTARY REFERENCES

1 Biely, P. et al. Mode of action of acetylxylan esterases on acetyl glucuronoxylan and acetylated oligosaccharides generated by a GH10 endoxylanase. *Biochim. Biophys. Acta* **1830**, 5075-5086 (2013).

2 Larsen, N. et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One* https://doi.org/10.1371/journal.pone.0009085 (2010).

3 Yu, Y., Lee, C., Kim, J. & Hwang, S. Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol. Bioeng.* **89**, 670-679 (2005).

4 Harvey, A. J., Hrmova, M., De Gori, R., Varghese, J. N. & Fincher, G. B. Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins* **41**, 257-269 (2000).

5 Shackelford, G. S., Regni, C. A. & Beamer, L. J. Evolutionary trace analysis of the alpha-D-phosphohexomutase superfamily. *Protein Sci.* **13**, 2130-2138 (2004).

6 Desai, M. S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**, 1339-1353 (2016).

Supplementary Data 1. List of putative CAZy domain-encoding genes upregulated on AcGGM and KGM. RNA-seq results of upregulation under β-mannans induction normalized to a glucose background are shown with Log_2-fold change values in bold type. Experiments were performed in three biological replicates. Locus tag numbers ROSINTL182_XXXXX are abbreviated with the last numbers after the hyphen. The presence of a predicted Signal Peptide (SP) was determined with SignalP (v.4.1). Upregulation of CAZyme genes associated with the degradation of other polysaccharides such as α-mannan, arabinan, α-rhamnosides and xyloglucan is ascribed to the presence of these components in the YCFA medium.

Locus Tag	CAZyme Family	Predicted Activity	Annotation	SP	Log_2-Fold Change AcGGM	KGM
07683	CBM23,CBM27,GH26	β-Mannanase	Fibronectin type III domain protein	Yes	**5.80**	**7.25**
05474	GH130	β-1,4-Mannooligosaccharide phosphorylase	Hypothetical protein	No	**5.63**	**5.42**
05475	GH130	β-1,4-Mannosylglucose phosphorylase	Hypothetical protein	No	**5.47**	**5.57**
05476	GT4	Epimerase	N-acylglucosamine 2-epimerase	No	**5.41**	**5.63**
08144	GH94	Cellobiose phosphorylase	N,N'-diacetylchitobiose phosphorylase	No	**4.48**	**3.58**
09616	GH23, GT61	Peptidoglycan lyase	NlpC/P60 family protein	No	**4.43**	**2.60**
08143	GH94	Cellobiose phosphorylase	N,N'-diacetylchitobiose phosphorylase	No	**4.31**	**3.76**
07684	GH3	β-Glucosidase	β-Glucosidase	No	**4.21**	**4.28**
05473	CE2	Acetyl xylan esterase	Hypothetical protein	No	**4.21**	**4.10**
07685	GH3	β-Glucosidase	β-Glucosidase	No	**4.13**	**4.32**
07158	GH23	Peptidoglycan lyase	Transglycosylase SLT domain protein	No	**4.13**	**2.99**
08193	GH43	α-L-Arabinofuranosidase	Arabinofuranosidase	No	**3.88**	**3.58**
06992	GH23	Peptidoglycan lyase	Putative lipoprotein	Yes	**3.83**	**2.18**
08195	CBM35,GH115	Xylan α-1,2-glucuronidase	Hypothetical protein	No	**3.74**	**3.40**
08196	GH8	Reducing-end-xylose releasing exo-oligoxylanase	Glycosyl hydrolase family 8	No	**3.69**	**3.54**
07032	CE1	Esterase	Hydrolase, alpha/beta domain protein	No	**3.69**	**1.81**
05471	CE2, CE3	Acetyl xylan esterase	GDSL-like protein	No	**3.55**	**3.47**
06494	CBM22, CBM9,GH10	Endo-1,4-β-xylanase	Endo-1,4- β -xylanase	No	**3.35**	**2.48**
08399	GH74	Endoglucanase	Candidate Xyloglucanase	No	**3.34**	**2.12**
06334	GH27	α-Galactosidase	α-Galactosidase	No	**3.33**	**2.38**
06340	GH51,GH43	β-Xylosidase/α-L-Arabinofuranosidase	Hypothetical protein	No	**3.27**	**1.80**
09628	GH3	β-Glucosidase	β-Glucosidase	No	**3.23**	**1.83**
05034	CBM35	Xylan-binding module	ABC Transporter	No	**3.19**	**1.87**
06343	CBM6, GH43	α-L-Arabinofuranosidase	Arabinofuranosidase	No	**3.07**	**1.98**
06338	CBM22, GH10	Endo-1,4-β-xylanase	Endo-1,4- β -xylanase	No	**3.06**	1.44
08567	GT41	UDP-GlcNAc:peptide Acetylglucosaminyltransferase	Tetratricopeptide repeat protein	Yes	**3.02**	**4.02**
09626	GH42, GH5	Cellulase/β-galactosidase	Hypothetical protein	No	**3.01**	**2.02**
08606	CBM50	Peptidoglycan-binding module	LysM domain protein	No	**2.85**	**2.06**
05969	GT2	Glycosyltransferase, group 2 family	Glycosyltransferase, group 2 family	No	**2.82**	0.68
09282	GT83	4-amino-4-deoxy-β-L-Arabinosyltransferase	Hypothetical protein	No	**2.76**	0.97

Locus Tag	CAZyme Family	Predicted Activity	Annotation	SP	Log_2-Fold Change AcGGM	Log_2-Fold Change KGM
06337	GH43	β-Xylosidase/α-L-Arabinofuranosidase	β-Xylosidase/α-L-Arabinofuranosidase	No	**2.75**	**1.54**
06342	CBM48, CE1	Acetyl xylan esterase	Putative xylanase	No	**2.67**	1.47
06341	GH43	α-L-Arabinofuranosidase	α-L-Arabinofuranosidase	No	**2.67**	**1.88**
07678	CBM61, GH53	endo-β-1,4-galactanase	Endo-β-1,4-galactanase	No	**2.65**	1.41
08358	CE10, CE7	Acetyl xylan esterase/arylesterase	Lysophospholipase	No	**2.62**	1.02
09560	GH78	α-L-Rhamnosidase	α-L-Rhamnosidase	No	**2.56**	1.35
05470[a]	GH1	β-Mannosidase/β-Glucosidase	Glycosyl hydrolase, family 1	No	**2.56**	**2.09**
06940	CBM37	Xylan-binding module	Repeat protein	No	**2.53**	1.17
06332	GH32, GH43	β-2,1-Fructosidase/α-L-Arabinofuranosidase	α-L-Arabinofuranosidase	No	**2.52**	1.41
08710	GH94	Cellobiose phosphorylase	Cellobiose phosphorylase	No	**2.50**	1.42
05469[a]	GH1	Mannose-6-phosphate isomerase	Mannose-6-phosphate isomerase	No	**2.38**	**2.15**
06339	GH10	Endo-1,4-β-xylanase	Endo-1,4-β-xylanase	Yes	**2.37**	0.51
05892	GH43	β-Xylosidase/α-L-Arabinofuranosidase	β-Xylosidase/α-L-Arabinofuranosidase	No	**2.32**	1.37
09562	GH94	Cellobiose phosphorylase	N,N'-diacetylchitobiose phosphorylase	No	**2.31**	1.17
08400	GH31	α-Glucosidase	α-Glucosidase	No	**2.27**	1.40
06335	GH43	β-Xylosidase/α-L-Arabinofuranosidase	β-Xylosidase/α-L-Arabinofuranosidase	No	**2.26**	0.87
05893	GH43	β-Xylosidase/α-L-Arabinofuranosidase	β-Xylosidase	No	**2.22**	1.36
05114	GH115	Xylan α-1,2-glucuronidase	Hypothetical protein	No	**2.19**	0.85
08711	GH94	Cellobiose phosphorylase	N,N'-diacetylchitobiose phosphorylase	No	**2.16**	0.97
09559	GH3	β-Glucosidase	β-Glucosidase	No	**2.15**	1.32
06336	GH95	α-L-Fucosidase	Hypothetical protein	No	**2.11**	1.00
09625	GH94	Cellobiose phosphorylase	N,N'-diacetylchitobiose phosphorylase	No	**2.09**	0.96
09492	GH2	β-Galactosidase	β-Galactosidase	No	**2.09**	1.03
09554	GH35	β-Galactosidase	β-Galactosidase	No	**1.99**	0.65
08556	GH5	Endo-β-1,4-glucanase/Cellulase	Endoglucanase A	No	**1.99**	**2.88**
09556	GH3	β-Glucosidase	β-Glucosidase	No	**1.92**	0.97
05638	GH35	β-Galactosidase	Glycosyl hydrolase, family 35	No	**1.86**	0.86
08720	GH38	α-Mannosidase	α-Mannosidase	No	**1.84**	0.69
05483	GH113	β-mannanase	Hypothetical protein	No	**1.80**	0.68
05574	GH119	α-Amylase	Hypothetical protein	No	**1.77**	0.74

Locus Tag	CAZyme Family	Predicted Activity	Annotation	SP	Log_2-Fold Change AcGGM	Log_2-Fold Change KGM
05891	CE12	Acetyl xylan esterase	Hypothetical protein	No	**1.77**	0.73
08203	GH42	β-Galactosidase/α-L-Arabinopyranosidase	β-Galactosidase	No	**1.77**	0.92
06331	GH43	β-Xylosidase/α-L-Arabinofuranosidase	β-Xylosidase/α-L-Arabinofuranosidase	No	**1.70**	0.84
09491	GH2	β-Galactosidase	β-Galactosidase	No	**1.70**	0.79
09478	GH73	Endo-β-N-acetylglucosaminidase	Hypothetical protein	No	**1.67**	0.54
05481	GH36	α-Galactosidase	α-Galactosidase	No	**1.64**	1.26
08825	GH13	α-Amylase	α-Amylase	No	**1.64**	0.84
08557	GH32, GH43	Fructan β-2,1-fructosidase/α-L-Arabinofuranosidase	Hypothetical protein	No	**1.63**	**2.51**
08719	GH2	β-Galactosidase	β-Galactosidase	No	**1.61**	0.58
07340	CE1	Esterase	Hydrolase	No	**1.60**	0.78
09481	GH24	Endo-β-N-acetylglucosaminidase	Hypothetical protein	No	**1.59**	0.47

[a]Loci ROSINTL182_05469 and ROSINTL182_05470 are annotated as two proteins of 522 and 232 amino acids, respectively. However, we corrected sequence ambiguities at the 3'-end of ROSINTL182_05470 and determined the absence of a stop codon in between the two genes.

Supplementary Data 2. List of *R. intestinalis* proteins upregulated when grown on AcGGM versus Glc. Proteins are annotated with the UniProt accession numbers and putative function. Locus tag numbers ROSINTL182_XXXXX are abbreviated with the last numbers after the hyphen. Values given for the carbon sources are Log_2 of MaxLFQ intensities, median of three biological replicates. Proteins involved in β-mannan catabolism are in bold.

			Median Log_2 LFQ Intensities			
UniProt ID	Locus tag	CAZy Family/Predicted function	Glc	AcGGM	Diff.[a]	Peptide count
C7GAP5	06975	Pyridine nucleotide-disulfide oxidoreductase	19.14	28.71	9.57	8
C7G6G5	**05478**	**MPP1, ABC transporter, permease protein**	**17.82**	**26.92**	**9.10**	**9**
C7GCP5	**07683**	**GH26 / β-mannosidase**	**20.69**	**29.25**	**8.56**	**85**
C7G6G2	**05475**	**GH130_1 / Mannosyl-glucose phosphorylase**	**21.76**	**29.58**	**7.82**	**41**
C7G6G4	**05477**	**MPP2, ABC transporter, permease protein**	**18.47**	**25.97**	**7.50**	**9**
C7G6G3	**05476**	**Mep, Epimerase**	**16.63**	**24.02**	**7.39**	**31**
C7G9B3	06492	GH1 / Arabinogalactan endo-galactanase	17.49	24.13	6.64	9
C7G6G1	**05474**	**GH130_2/ Mannoside phosphorylase**	**23.91**	**30.17**	**6.26**	**28**
C7G6G6	**05479**	**ABC transporter, mannan-binding protein**	**24.84**	**30.83**	**5.98**	**21**
C7GE35	08187	Hypothetical protein	21.07	27.01	5.95	40
C7GCP6	**07684**	**GH3B / β -glucosidase**	**19.37**	**25.19**	**5.82**	**32**
C7GCP7	**07685**	**GH3A / β -glucosidase**	**20.70**	**26.45**	**5.75**	**48**
C7GE41	08193	GH43 / α-arabinofuranosidase	18.17	23.82	5.65	28
C7GF65	08570	ABC transporter, solute-binding protein	18.87	24.44	5.56	10
C7GF66	08571	Hypothetical protein	18.33	23.89	5.56	17
C7GER5	08417	Maturation rSAM protein, HydE	17.50	22.76	5.26	4
C7GF51	08556	GH1 / Endoglucanase	18.59	23.57	4.97	5
C7G6G0	**05473**	**Carbohydrate esterase, CE2**	**19.06**	**24.02**	**4.96**	**16**
C7GF52	08557	Hypothetical protein	19.32	24.18	4.86	18
C7GGA9	08977	Sugar-binding domain protein	18.78	23.58	4.80	5
C7G6F8	**05471**	**Carbohydrate esterase, CEX**	**21.81**	**26.53**	**4.72**	**29**
C7GE44	08196	GH8 /reducing-end xylose releasing oligoxylanase	18.12	22.83	4.71	12
C7GE47	08199	ABC transporter, xylan-binding protein	21.72	26.36	4.64	27
C7GE46	08198	ABC transporter, permease protein	19.22	23.84	4.62	2
C7GDZ0	08142	Transcriptional regulator, LacI family	19.64	24.12	4.49	12
C7G8A2	06126	Enolase	13.92	18.33	4.41	30
C7GB09	07093	Riboflavin biosynthesis protein, RibH	20.58	24.99	4.41	15
C7GE51	08203	β-galactosidase	20.23	24.63	4.40	23
C7GCG9	07606	Hypothetical protein	18.15	22.53	4.38	4
C7G6H0	**05483**	**GH113**	**19.57**	**23.93**	**4.36**	**13**
C7G9G1	06540	Hypothetical protein	19.35	23.39	4.04	7
C7GF55	08560	Hypothetical protein	23.21	27.24	4.03	14
C7GD06	07806	30S ribosomal protein S10	21.10	25.02	3.92	7
C7GCL0	07648	Membrane protein insertase, YidC/Oxa1 family	16.19	20.07	3.88	14
C7G9Z7	06727	TRAP transporter solute receptor, DctP family	19.40	23.25	3.85	8
C7G830	06054	30S ribosomal protein	19.86	23.66	3.80	4
C7GE73	08224	Endonuclease III	16.77	20.53	3.76	4
C7GEV7	08459	Signal recognition particle protein	17.84	21.59	3.75	29
C7GH91	09312	Putative translation elongation factor G	20.71	24.45	3.74	41

UniProt ID	Locus tag	CAZy Family/Predicted function	Median Log_2 LFQ Intensities		Diff.[a]	Peptide count
			Glc	AcGGM		
C7G5M9	05187	IS66 family element, transposase	17.74	21.47	3.73	9
C7GHZ4	09566	β-lactamase	19.45	23.11	3.66	9
C7GF62	08567	Tetratricopeptide repeat protein	18.75	22.34	3.59	5
C7GFK0	08707	Hypothetical protein	21.04	24.62	3.58	7
C7G9U2	06672	GH38 / α-mannosidase	20.98	24.51	3.54	42
C7GDY7	08139	Hypothetical protein	21.74	25.27	3.53	9
C7G7D3	05800	ABC transporter, ATP-binding protein	18.03	21.52	3.49	27
C7GB06	07090	Riboflavin biosynthesis protein, RibD	17.66	21.12	3.46	8
C7GD03	07803	50S ribosomal protein L23	21.22	24.67	3.45	8
C7G702	05669	RelA/SpoT family protein	21.79	25.15	3.35	32
C7GCZ6	07796	30S ribosomal protein S17	21.65	24.98	3.33	8
C7GD85	07886	ABC transporter, ATP-binding protein	19.08	22.39	3.32	41
C7GCG3	07600	ABC transporter, solute-binding protein	20.18	23.44	3.26	16
C7G7L4	05888	ABC transporter, glycerol-3-phosphate transport system	18.66	21.89	3.24	21
C7G6G9	**05482**	**Pgm, Phosphomannomutase**	**21.70**	**24.84**	**3.14**	**62**
C7GDL8	08019	Transketolase, Pentose phosphate pathway	18.99	22.12	3.14	4
C7GE45	08197	ABC transporter, permease protein	19.66	22.78	3.12	6
C7GB02	07086	Response regulator receiver domain protein	18.91	21.97	3.06	10
C7GDY8	08140	Hypothetical protein	22.35	25.39	3.04	3
C7GDI0	07981	Isoprenoid synthesis protein, IspH	18.51	21.54	3.03	6
C7GDW1	08113	Hsp20/alpha crystallin family protein	17.12	20.12	3.00	3
C7GF11	08514	Hypothetical protein	17.75	20.75	3.00	4
C7G6K9	05523	Polyphosphate:AMP phosphotransferase	20.68	23.66	2.98	17
C7GBV3	07388	Flagellar basal body protein	21.61	24.58	2.97	22
C7GCJ0	07628	Pantothenate synthetase	23.69	26.66	2.96	33
C7GDZ1	08143	Glycosyltransferase family 36	22.95	25.91	2.96	18
C7G5E4	05102	Aminopeptidase	20.95	23.91	2.95	38
C7GHF0	09371	GH31 / putative α-xylosidase	21.55	24.50	2.95	21
C7GCI9	07627	Aspartate 1-decarboxylase	21.25	24.17	2.92	7
C7G750	05717	Hypothetical protein	18.61	21.51	2.90	6
C7GDM1	08023	L-arabinose isomerase	18.61	21.51	2.90	13
C7G674	05384	Putative guanine deaminase	19.64	22.53	2.90	18
C7G5X1	05280	VanW-like protein	21.63	24.50	2.87	12
C7GE21	08173	DNA helicase	20.65	23.51	2.86	37
C7G7T5	05959	Hydrolase	17.63	20.49	2.86	8
C7G5L1	05169	Malonyl CoA-acyl carrier protein transacylase	19.33	22.19	2.86	26
C7GFZ7	08854	Translation initiation factor IF-3	22.48	25.25	2.77	11
C7GCJ1	07629	3-methyl-2-oxobutanoate hydroxymethyltransferase	21.62	24.38	2.77	22
C7GDL7	08018	Transketolase, Pentose phosphate pathway	21.22	23.98	2.76	5
C7G9C2	06501	Hypothetical α-amylase	23.76	26.51	2.76	21
C7GDN0	08031	Hypothetical protein	23.37	26.10	2.73	12
C7GDZ2	08144	Hypothetical protein	24.69	27.42	2.73	45
C7G6R3	05578	Hypothetical protein	20.52	23.25	2.72	5
C7GCL9	07657	ParB-like protein	18.18	20.88	2.71	15
C7GEV6	08458	30S ribosomal protein S16	20.87	23.58	2.70	5

UniProt ID	Locus tag	CAZy Family/Predicted function	Median Log_2 LFQ Intensities		Diff.[a]	Peptide count
			Glc	AcGGM		
C7G6Y0	05647	Putative catabolite control protein A	17.98	20.68	2.70	3
C7G6F7	**05470**	**GH1**	**20.38**	**23.03**	**2.65**	**16**
C7G7Q2	05926	Ribosomal protein S12 methylthiotransferase, RimO	21.85	24.49	2.65	21
C7GEF2	08304	von Willebrand factor type A domain protein	22.12	24.77	2.65	17
C7GE38	08190	ABC transporter, solute-binding protein	22.38	25.01	2.63	12
C7G8L8	06244	Hypothetical phosphoketolase	21.25	23.86	2.60	42
C7G7C5	05792	ABC transporter, ATP-binding protein	18.52	21.11	2.59	16
C7GDS7	08079	EDD domain protein, DegV family	22.43	24.99	2.56	13
C7GCB6	07551	ABC transporter, ATP-binding protein	23.30	25.82	2.53	21
C7GE37	08189	ABC transporter, permease protein	18.96	21.48	2.52	3
C7GCF5	07592	Site-determining protein	23.12	25.64	2.52	18
C7GC93	07528	Phosphopentomutase	19.20	21.71	2.51	20
C7GE43	08195	GH115 / α-glucuronidase	18.06	20.56	2.50	9
C7GDL9	08020	Transketolase, thiamine diphosphate binding domain protein	21.69	24.19	2.50	5
C7GEF1	08303	Hypothetical protein	26.54	29.02	2.48	74
C7G9L8	06597	Cell shape determining protein, MreB/Mrl family	21.14	23.61	2.48	20
C7GAG5	06895	Hypothetical protein	20.29	22.74	2.45	10
C7G8E5	06169	RNA pseudouridine synthase	18.41	20.83	2.42	4
C7GCX7	07777	30S ribosomal protein S11	22.91	25.31	2.40	12
C7GDX0	08122	HD domain protein	18.21	20.61	2.40	11
C7G9H0	06549	Glycerophosphotransferase	18.30	20.69	2.39	41
C7GGC3	08991	TOBE domain protein	24.99	27.38	2.38	3
C7G5C9	05087	Hypothetical protein	18.30	20.67	2.38	4
C7GDQ8	08059	Hypothetical protein	23.67	26.04	2.37	10
C7GGG3	09031	Signal peptidase I	17.66	20.03	2.37	5
C7GBY2	07417	ABC transporter, solute-binding protein	21.47	23.82	2.35	12
C7G9R9	06648	Dipeptidase	16.74	19.09	2.35	16
C7G7B6	05783	Hypothetical protein	18.06	20.41	2.34	8
C7GG32	08894	Putative glycogen debranching enzyme	17.64	19.96	2.33	22
C7GE90	08242	DNA mismatch repair protein, MutS	20.01	22.33	2.32	14
C7GDQ7	08058	Hypothetical protein	22.86	25.18	2.31	33
C7GCZ5	07795	50S ribosomal protein L14	24.19	26.49	2.30	13
C7GA10	06762	Cell Wall Hydrolase	23.34	25.64	2.30	10
C7GDM0	08021	Putative transaldolase	21.84	24.13	2.29	19
C7GDH7	07978	Leucine Rich Repeat protein	24.43	26.71	2.29	65
C7GC36	07471	Siphovirus Gp157	19.46	21.74	2.28	7
C7GEC8	08280	Hypothetical protein	20.47	22.74	2.27	11
C7GCF7	07594	Aminotransferase, class I/II	18.27	20.54	2.27	20
C7G7D2	05799	ABC transporter, ATP-binding protein	19.97	22.21	2.24	37
C7GD01	07801	30S ribosomal protein S19	25.09	27.32	2.23	16
C7GE34	08186	Hypothetical protein	19.42	21.64	2.22	4
C7G7X0	05994	POTRA domain protein, FtsQ-type	17.38	19.61	2.22	2
C7G8I6	06212	Diguanylate cyclase domain protein	19.56	21.78	2.22	9
C7GF48	08551	tRNA N6-adenosine threonylcarbamoyltransferase	20.09	22.30	2.22	5
C7GEE4	08296	Hypothetical protein	21.66	23.87	2.22	4

UniProt ID	Locus tag	CAZy Family/Predicted function	Median Log_2 LFQ Intensities		Diff.[a]	Peptide count
			Glc	AcGGM		
C7GHX4	09545	ATP-dependent metallopeptidase HflB	18.40	20.61	2.21	10
C7GAT0	07010	Glycerol-3-phosphate responsive antiterminator	20.58	22.77	2.19	12
C7GDH3	07974	Electron transport complex subunit G	18.83	21.00	2.17	7
C7GEV1	08453	50S ribosomal protein L19	24.84	27.00	2.16	16
C7GCE8	07585	Uridine kinase	18.81	20.95	2.14	7
C7G5B8	05076	Penicillin-binding protein	16.98	19.12	2.14	30
C7GCZ4	07794	50S ribosomal protein L24	18.24	20.37	2.13	3
C7G6G7	**05480**	**Transcriptional regulator, LacI family**	**21.38**	**23.49**	**2.11**	**9**
C7G956	06433	Transcriptional regulator, MarR family	18.39	20.49	2.10	2
C7GE31	08183	GH53 / Arabinogalactan β-galactanase	21.37	23.46	2.09	7
C7GE96	08248	Cell shape determining protein, MreB/Mrl family	20.38	22.45	2.07	16
C7GBN0	07315	ABC transporter, substrate-binding protein, family 5	17.55	19.61	2.06	24
C7GCY9	07789	50S ribosomal protein L6	25.88	27.92	2.05	19
C7GCL6	07654	Putative sensor protein, DegS	18.71	20.73	2.02	8
C7GAP6	06976	FAD dependent oxidoreductase	18.83	20.84	2.01	4
C7GGY8	09209	Hypothetical protein	28.44	30.44	2.00	10
C7GED6	08288	Hypothetical protein	20.37	22.36	1.99	13
C7G9N1	06610	Hypothetical protein	19.55	21.53	1.99	4
C7GAG4	06894	Hypothetical protein	21.57	23.49	1.92	11
C7GCL5	07653	Hypothetical protein	20.65	22.54	1.89	22
C7G890	06114	Hypothetical protein	19.63	21.51	1.88	18
C7GBN7	07322	Oxidoreductase, aldo/keto reductase family protein	19.49	21.37	1.87	15
C7GE32	08184	GH2 / β-galactosidase	21.72	23.57	1.85	42
C7GAY7	07068	Hypothetical protein	22.70	24.52	1.81	2
C7G6G8	**05481**	**GH36 / α-galactosidase**	**23.85**	**25.62**	**1.77**	**57**
C7GCF2	07589	Thiamine biosynthesis pathway, ThiH	20.16	21.92	1.77	4
C7G5Q2	05210	SPFH/Band 7/PHB domain protein	19.95	21.70	1.75	28
C7GDL6	08017	Rribulose-5-phosphate 4-epimerase	22.71	24.44	1.73	9
C7G7K2	05876	Sporulation protein YtfJ	19.72	21.39	1.68	8
C7GDX7	08129	Tetratricopeptide repeat protein	20.78	22.46	1.67	20
C7GDM3	08024	Carbohydrate kinase	19.94	21.57	1.64	7
C7G8K4	06230	Cobalt transport protein	17.60	19.23	1.63	9
C7GDH8	07979	CheR methyltransferase, glutamate biosynthesis	20.36	21.91	1.56	31
C7GE91	08243	Thymidylate synthase, ThyX	20.78	22.33	1.54	21

[a]Difference between AcGGM and Glc Log_2-fold values.

Discussion and perspectives

The interest in HGM research has increased tremendously in recent years, and it is generally accepted that health benefits from the intake of dietary fibers and prebiotics to a large extent can be coupled to changes in HGM composition and catabolic output. Focus has been directed towards the SCFAs products of fiber fermentation, due to their observed positive effects on immune and metabolic homeostasis and relevance in different diseases. However, there is limited insight into how to modify the HGM composition and we are lacking knowledge about individual species as well as communities in regards to dietary fiber utilization. In this thesis the molecular apparatus for utilization of two selected dietary fibers, xylan and mannan, are characterized for the butyrate producing *R. intestinalis.* The studies presented here are the first to describe utilization of these two glycans by Firmicutes from the HGM and can thus potentially serve as a model system.

Impact of glycan utilization and transport

Modular cell-attached GHs are identified to be responsible for the initial capture of complex xylans and mannans by *R. intestinalis.* GHs from *Clostridium* cluster XIVa of the HGM have a variety of binding modules and substrates, while seemingly performing a similar function; capture and hydrolysis of glycans into oligosaccharides at the cell surface prior to import facilitated by one or more ABC-transporters[157], [158]. Multiplicity and variability of CBMs seem to be a common feature of extracellular enzymes from butyrate-producing Firmicutes from the human gut niche[157], [159] and can be an advantage in capture and prolonged contact of enzymes with glycans. *R. intestinalis* was shown to be associated to insoluble xylans, including wheat bran, whereas Bacteriodes spp. were enriched in solubilized xylan fractions[61], [160], which may be facilitated by the multiple CBMs.

In order to understand the selectiveness of the top upregulated ABC-transporters in the two studies, binding of oligosaccharides to the SBP components was characterized. The specificity of SBPs is typically used as a proxy for the specificity of the ABC transporter system. However, TMDs also contribute to the specificity, but lack sufficient affinity, [133],. Studies of the maltose uptake system in *E. coli*, have shown that TMDs possess a substrate binding pocket and display specificity for maltose in the absence of the SBP, but the apparent K_{m} was significantly increased[161]. This suggests that SBP is crucial

for high affinity capture, which defines the specificity of the transporter during substrate limiting conditions, e.g. in the human gut. The high conservation of TMD sequences for related substrates, compared to high divergence in the sequences of SBP, is consistent with the specificity determining role of SBPs. Moreover, the uptake profiles during growth of bacteria on oligosaccharides, reflect affinity profiles of the SBP component [162]. Generating gene knockouts of transporter genes is powerful in confirming their roles, but this is currently not feasible in *Roseburia* species.

For primary degraders that possess extracellular GHs, the import of oligosaccharides is closely coupled to the hydrolysis of glycans at the cell surface. Firmicutes import oligosaccharides hydrolyzed from polymers by modular cell-attached GHs and the complete depolymerisation is performed by intracellular GHs and accessory CAZymes. In contrast, Bacteriodetes (mainly from the *Bacteroides* genus) rely more on extracellular degradation by one or several GHs and transport of larger oligosaccharides into the periplasm for subsequent degradation into monosaccharides, which are transported over the inner membrane via MFS transporters[94], [163]. Instead of having multiple CBMs in extracellular GHs, Bacteriodetes often have fewer or no CBMs in their GHs and rely on SGBPs that are a part of the Sus-like system, for capture of substrates. Common to both phyla is that only partial degradation of glycans occurs extracellularly. The extent of alignment between the degradation products and the preference of uptake systems in one organism is likely to govern how selfish or sharing the glycan utilization strategy will be. If the degradation products are fully fit to the uptake systems, then very little leakage is likely to occur, whereas the generation of a broader set of glycans from extracellular GHs is likely to enable taxa possessing efficient uptake systems to compete for some of the degradation productions.

To facilitate a process where minimum amounts of oligosaccharides are lost to the surrounding environment, the surface GHs must operate at a rate that minimizes accumulation of oligomers at the surface. This form of metabolism is deployed by *B. thetaiotaomicron* during the utilization of yeast α-mannans, in which the generated manno-oligosaccharides are rapidly imported into the periplasm, conferring no direct benefits on neighboring species[163]. In the growth experiment with *R. intestinalis* no oligosaccharides were detected in the supernatants, suggesting a similar efficient uptake by the respective ABC-transporters[63], [66].

Cross feeding

Cross feeding between HGM is observed on fermentation products, as described for *R. intestinalis* in section 1.3. The dynamics of the HGM community are influenced by this type of cross feeding and it has been investigated in a well-defined small representative gut microbial community consisting of the three strains *B. hydrogenotrophica, F. prausnitzii* and *R. intestinalis*[164]. In this study, it was shown that strains grown in co-cultures behaved differently than those in monocultures, which was confirmed at the transcriptional level. Therefore, studying bacteria in mono-cultures can be insufficient to understand the complex competitive environment of our gut. Mono-cultures are, however, increasingly carried out to obtain data about individual members of the HGM. Such insight, as generated in this thesis, is important, but it is at least as important to keep in mind that the individual bacterium may behave different in a community.

An example of changed expression is the study where *B. thetaiotaomicron* and *E. rectale* were co-colonized in mice[165]. *E. rectale* adapted to the co-culture by downregulating many of its GHs for glycan utilization and instead upregulated oligosaccharide transporters, while *B. thetaiotaomicron* upregulated polysaccharide utilization loci encoding numerous GHs. In our experiments the expression of the GH10 did not appear to be downregulated when *R. intestinalis* was grown in a co-culture with *B. ovatus*, however, caution needs to be taken in inferring general models from single studies. To further investigate this, one could do a similar *in vivo* experiment in mice as described above, but with a defined mini-microbiome and examine the transcriptional profile of *R. intestinalis* in an environment mirroring our gut.

Regulation

Plant derived diets contain more than one type of dietary fibers and the utilization of the different glycans requires activation of multiple CAZymes and transporters. Many HGM species have the ability to utilize a variety of glycans, but they do typically not target all simultaneously. Instead, bacteria have evolved a hierarchical utilization of glycans through a phenomenon that resembles catabolite repression[166]. An experiment with *B. thetaiotaomicron* showed that when the bacteria was presented to mucin glycans and other carbohydrates, then the degradation of mucin glycans was consistently repressed[167]. In the same study, it was also shown that this metabolic hierarchy is species specific, with different species exhibiting variable and sometimes opposite rank

orders for some glycans. Structural aspects of glycan substrates have in addition been demonstrated to affect prioritization, including linkage profiles, conformation, and molecular size[166]. This type of metabolic hierarchy has mainly been described for Bacteriodetes, however, Firmicutes - with the ability to utilize several glycans - have most likely evolved a similar approach for prioritizing. The presence of typically multiple transcriptional regulators in glycan utilization gene clusters of Firmicutes is supportive of the importance of a complex catabolic regulation.

Operon specific transcriptional regulators enable bacteria to adjust their carbohydrate catabolism, to ensure that cellular energy is conserved[168]. Regulatory proteins of the LacI-GalR family are often associated with glycan utilization genes in Gram positive bacteria, in order to induce or repress the genes in a substrate-specific manner[169]. The identified xylan and mannan utilization gene clusters are predicted to encode LacI-type transcriptional regulators. These LacI genes; ROSINTL182_08194 and ROSINTL102_05480, are regulated at the transcriptional level by induction of xylan and mannan, respectively, suggesting that they are important in transcriptional regulation of specific genes in the presence of these two glycans. We did not pursue LacI genes in our work, although further investigation could establish if the two LacI genes directly interact with specific operators of the promotor regions and which regulatory signals (substrates) act on them.

Discovery of novel CAZy families

Xylan and mannan are, as described in section 1.7-8, usually acetylated at C2 and C3. Thus, a molecular machinery that mediates complete breakdown of these glycan requires efficient CEs for deacetylation. In this thesis, three CE were characterized. *Ri*CEX, *Ri*CE2 and *Ri*AXE. The latter two show 18.6% identity to each other (Appendix 1) and share conserved residues with the SGNH lipase esterase superfamily (Pfamcd00229). However, both CEs lack functionally characterized homologues, suggesting that they belong to new families of CEs. *Ri*AXE is efficient for deacetylation of both C2 and C3 in xylan, but with a preference for C2 decorations[63]. *Ri*CEX hydrolyzed in concert with *Ri*CE2 almost complete deacetylation of mannan, but seems to have a preference for oligosaccharides with a degree of acetyl substitution ≥ 2[66]. To pursue the potential of *Ri*AXE belonging to a new family, we sought to solve the crystal structure, but we did not succeed in obtaining crystals. To get further insight on these two potential new CE families, solving the structures of either enzyme would be valuable.

CBMs of cell-attached enzymes are important for the recognition of accessible glycans in the complex gut environment. The modular cell-attached xylanase *Ri*Xyn10A has four CBM modules to facilitate binding to xylan. We have characterized one of these modules, CBMx, and proposed the assignment into a novel CBM family[63], [170], chapter 4. CBMx lacks functionally described homologues, but related sequences from Firmicutes *Clostridium* XIVa members are available, suggesting this novel family is associated or enriched within Clostridiales. Characterization of the binding properties showed that it is selective for xylan and with the affinity increasing for oligosaccharides from DP4-DP6, with a K_D=490 µM for X6, suggesting that CBMx is a low affinity binding CBM. The crystal structure and the NMR titration highlight preference to arabino-decordated xylan as compared to glucuronodecorated xylan. Further studies are needed to evaluate crucial residues in the binding site and the contribution of this CBM in the context of the full-length xylanase.

Conclusion

The present Ph.D. thesis summarizes my research on the butyrate producing *R. intestinalis* from the HGM that focused on bringing new insight to the molecular mechanisms for xylan and mannan utilization by Firmicutes. The work has resulted in the identification and characterization of key enzymes and transporters. This has together with data from the growth experiments provided a detailed view of how *R. intestinalis* utilize complex glycans in a competitive environment with other primary glycan degraders. The results have highlighted that the differentiation in glycan capture and uptake preferences might contribute the ability to co-growth on dietary fibers by different taxa - future investigations are needed to validate the generality of this finding. The studies have also enabled us discover two CEs and one CBM belonging to novel CAZy families, with the crystal structure of the CBMx further highlighting the uniqueness of this CBM, as compared to its xylan specific counterparts.

Our findings are the first to describe the utilization of xylan and mannan by *R. intestinalis* and can potentially serve as a model system for utilization of other glycans by this taxonomic group of Firmicutes from the HGM. Finally, the findings contribute the understanding of the interaction between diet and the HGM. This might be exploited to selectively promote key members of the healthy HGM through a prebiotic or synbiotic approach, which can serve as a way to enhance local production on butyrate in the colon, thereby taken advantage of butyrate's health promoting effects.

References

[1] L. E. Tailford, E. H. Crost, D. Kavanaugh, and N. Juge, "Mucin glycan foraging in the human gut microbiome," *Front. Genet.*, vol. 5, no. FEB, 2015.

[2] G. P. Donaldson, S. M. Lee, and S. K. Mazmanian, "Gut biogeography of the bacterial microbiota," *Nat. Rev. Microbiol.*, vol. 14, no. 1, pp. 20–32, 2015.

[3] R. Sender, S. Fuchs, and R. Milo, "Revised Estimates for the Number of Human and Bacteria Cells in the Body," *PLoS Biol.*, vol. 14, no. 8, pp. 1–14, 2016.

[4] H. P. Browne, B. A. Neville, S. C. Forster, and T. D. Lawley, "Transmission of the gut microbiota: spreading of health," *Nat. Rev. Microbiol.*, vol. 15, no. 9, pp. 531–543, 2017.

[5] J. Qin *et al.*, "A human gut microbial gene catalog established by metagenomic sequencing," *Nature*, vol. 464, no. 7285, pp. 59–65, 2010.

[6] G. M. Nava, H. J. Friedrichsen, and T. S. Stappenbeck, "Spatial organization of intestinal microbiota in the mouse ascending colon," *ISME J.*, vol. 5, no. 4, pp. 627–638, 2011.

[7] M. J. Bull and N. T. Plummer, "Part 1: The Human Gut Microbiome in Health and Disease.," *Integr. Med. (Encinitas).*, vol. 13, no. 6, pp. 17–22, 2014.

[8] J. M. Rodríguez *et al.*, "The composition of the gut microbiota throughout life, with an emphasis on early life," *Microb. Ecol. Heal. Dis.*, vol. 26, no. 0, pp. 1–17, 2015.

[9] J. E. Koenig *et al.*, "Succession of microbial consortia in the developing infant gut microbiome," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 108, no. Suppl 1, pp. 4578–4585, Mar. 2011.

[10] C. N. de Almada, C. Nunes de Almada, R. C. R. Martinez, and A. de S. Sant'Ana, "Characterization of the intestinal microbiota and its interaction with probiotics and health impacts," *Appl. Microbiol. Biotechnol.*, vol. 99, no. 10, pp. 4175–4199, 2015.

[11] J. L. Sonnenburg and F. Bäckhed, "Diet–microbiota interactions as moderators of human metabolism," *Nature*, vol. 535, no. 7610, pp. 56–64, 2016.

[12] C. De Filippo *et al.*, "Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa," *Proc. Natl. Acad. Sci.*, vol. 107, no. 33, pp. 14691–14696, 2010.

[13] D. Rothschild *et al.*, "Environment dominates over host genetics in shaping human gut microbiota," *Nature*, vol. 555, p. 210, Feb. 2018.

[14] B. Wang, M. Yao, L. Lv, Z. Ling, and L. Li, "The Human Microbiota in Health and Disease," *Engineering*, vol. 3, no. 1, pp. 71–82, Feb. 2017.

[15] A. J. Baumler and V. Sperandio, "Interactions between the microbiota and pathogenic bacteria in the gut," *Nature*, vol. 535, no. 7610, pp. 85–93, Jul. 2016.

[16] T. Gensollen, S. S. Iyer, D. L. Kasper, and R. S. Blumberg, "How colonization by microbiota in early life shapes the immune system," *Science*, vol. 352, no. 6285, pp. 539–544, Apr. 2016.

[17] P. D. Cani and N. M. Delzenne, "Gut microflora as a target for energy and metabolic homeostasis," *Curr. Opin. Clin. Nutr. Metab. Care*, vol. 10, no. 6, 2007.

[18] C. Manichanh *et al.*, "Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach," *Gut*, vol. 55, no. 2, pp. 205–211, Feb. 2006.

[19] T. R. Abrahamsson, H. E. Jakobsson, A. F. Andersson, B. Björkstén, L. Engstrand, and M. C. Jenmalm, "Low gut microbiota diversity in early infancy precedes asthma at school age," *Clin. Exp. Allergy*, vol. 44, no. 6, pp. 842–850, Jun. 2014.

[20] R. E. Ley, P. J. Turnbaugh, S. Klein, and J. I. Gordon, "Microbial ecology: Human gut microbes associated with obesity," *Nature*, vol. 444, no. 7122, pp. 1022–1023, Dec. 2006.

[21] A. Koliada *et al.*, "Association between body mass index and Firmicutes/Bacteroidetes ratio in an adult Ukrainian population," *BMC Microbiol.*, vol. 17, p. 120, May 2017.

[22] E. Le Chatelier *et al.*, "Richness of human gut microbiome correlates with metabolic markers," *Nature*, vol. 500, p. 541, Aug. 2013.

[23] I. Martínez *et al.*, "The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes," *Cell Rep.*, vol. 11, no. 4, pp. 527–538, Nov. 2015.

[24] S. L. Schnorr *et al.*, "Gut microbiome of the Hadza hunter-gatherers," *Nat. Commun.*, vol. 5, p. 3654, 2014.

[25] W. S. Garrett *et al.*, "Communicable Ulcerative Colitis Induced by T-bet Deficiency in the Innate Immune System," *Cell*, vol. 131, no. 1, pp. 33–45, Nov. 2007.

[26] M. Vijay-Kumar *et al.*, "Metabolic Syndrome and Altered Gut Microbiota in Mice Lacking Toll-Like Receptor 5," *Science*, vol. 328, no. 5975, pp. 228–231, Apr. 2010.

[27] A. Spor, O. Koren, and R. Ley, "Unravelling the effects of the environment and host genotype on the gut microbiome," *Nat. Rev. Microbiol.*, vol. 9, no. 4, pp. 279–290, 2011.

[28] P. Louis and H. J. Flint, "Formation of propionate and butyrate by the human colonic microbiota," *Environ. Microbiol.*, vol. 19, no. 1, pp. 29–41, 2017.

[29] G. den Besten, K. van Eunen, A. K. Groen, K. Venema, D.-J. Reijngoud, and B. M. Bakker, "The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism," *J. Lipid Res.*, vol. 54, no. 9, pp. 2325–2340, 2013.

[30] D. Ríos-Covián, P. Ruas-Madiedo, A. Margolles, M. Gueimonde, C. G. De los Reyes-Gavilán, and N. Salazar, "Intestinal short chain fatty acids and their link with diet and human health," *Front. Microbiol.*, vol. 7, no. FEB, pp. 1–9, 2016.

[31] J. H. Cummings, E. W. Pomare, W. J. Branch, C. P. Naylor, and G. T. Macfarlane, "Short chain fatty acids in human large intestine, portal, hepatic and venous blood.," *Gut*, vol. 28, no. 10, pp. 1221–1227, Oct. 1987.

[32] P. Louis, G. L. Hold, and H. J. Flint, "The gut microbiota, bacterial metabolites and colorectal cancer," *Nat. Rev. Microbiol.*, vol. 12, no. 10, pp. 661–672, 2014.

[33] P. Louis, H. J. Flint, and C. Michel, "How to Manipulate the Microbiota: Prebiotics," *Adv. Exp. Med. Biol*, vol. 902, pp. 119–142, 2016.

[34] S. Anand, H. Kaur, and S. S. Mande, "Comparative *in silico* analysis of butyrate production pathways in gut commensals and pathogens," *Front. Microbiol.*, vol. 7, no. December, pp. 1–12, 2016.

[35] P. Louis and H. J. Flint, "Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine," *FEMS Microbiol. Lett.*, vol. 294, no. 1, pp. 1–8, 2009.

[36] A. Koh, F. De Vadder, P. Kovatcheva-Datchary, and F. Bäckhed, "From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites," *Cell*, vol. 165, no. 6, pp. 1332–1345, 2016.

[37] A. Rivière, M. Selak, D. Lantin, F. Leroy, and L. De Vuyst, "Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut," *Front. Microbiol.*, vol. 7, no. JUN, 2016.

[38] D. J. Morrison and T. Preston, "Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism," *Gut Microbes*, vol. 7, no. 3, pp. 189–200, 2016.

[39] D. R. Donohoe *et al.*, "The warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation," *Mol. Cell*, vol. 48, no. 4, pp. 612–626, Nov. 2012.

[40] M. Pozuelo *et al.*, "Reduction of butyrate- and methane-producing microorganisms in patients with Irritable Bowel Syndrome," *Sci. Rep.*, vol. 5, no. 1, p. 12693, 2015.

[41] R. Kumari, V. Ahuja, and J. Paul, "Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India," *World J. Gastroenterol.*, vol. 19, no. 22, pp. 3404–3414, Jun. 2013.

[42] N. Wu *et al.*, "Dysbiosis Signature of Fecal Microbiota in Colorectal Cancer Patients," *Microb. Ecol.*, vol. 66, no. 2, pp. 462–470, 2013.

[43] K. Takahashi *et al.*, "Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease," *Digestion*, vol. 93, no. 1, pp. 59–65, 2016.

[44] M. Arumugam *et al.*, "Enterotypes of the human gut microbiome," *Nature*, vol. 473, no. 7346, pp. 174–180, 2011.

[45] F. Yasmin *et al.*, "Cesarean Section, Formula Feeding, and Infant Antibiotic Exposure: Separate and Combined Impacts on Gut Microbial Changes in Later Infancy," *Front. Pediatr.*, vol. 5, no. September, p. 200, 2017.

[46] P. O. Sheridan *et al.*, "Polysaccharide utilisation loci and nutritional specialisation in a dominant group of butyrate-producing human colonic Firmicutes.," *Microb. Genomics*, vol. 2, no. August 2015, 2016.

[47] A. M. Mowat and W. W. Agace, "Regional specialization within the intestinal immune system," vol. 14, p. 667, Sep. 2014.

[48] P. Van den Abbeele *et al.*, "Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an *in vitro* gut model.," *ISME J.*, vol. 7, no. 5, pp. 949–61, 2013.

[49] J. Qin *et al.*, "A metagenome-wide association study of gut microbiota in type 2 diabetes," *Nature*, vol. 490, no. 7418, pp. 55–60, 2012.

[50] A. R. Erickson *et al.*, "Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease," *PLoS One*, vol. 7, no. 11, 2012.

[51] B. P. Willing *et al.*, "A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes.," *Gastroenterology*, vol. 139, no. 6, p. 1844–1854.e1, Dec. 2010.

[52] Z. Shen *et al.*, "Insights into *Roseburia intestinalis* which alleviates experimental colitis pathology by inducing anti-inflammatory responses," *J. Gastroenterol. Hepatol.*, vol. 33, no. 10, pp. 1751–1760, Oct. 2018.

[53] T. Wang *et al.*, "Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers," *ISME J.*, vol. 6, no. 2, pp. 320–329, Feb. 2012.

[54] Z. Tamanai-Shacoori *et al.*, "*Roseburia* spp.: a marker of health?," *Future Microbiol.*, vol. 12, no. 2, pp. 157–170, 2017.

[55] K. Kasahara *et al.*, "Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model," *Nat. Microbiol.*, vol. 3, no. 12, pp. 1461–1471, 2018.

[56] S. H. Duncan, R. I. Aminov, K. P. Scott, P. Louis, T. B. Stanton, and H. J. Flint, "Proposal of Roseburia faecis sp. nov., Roseburia hominis sp. nov. and Roseburia inulinivorans sp. nov., based on isolates from human faeces," *Int. J. Syst. Evol. Microbiol.*, vol. 56, no. 10, pp. 2437–2441, 2006.

[57] S. H. Duncan, G. L. Hold, A. Barcenilla, C. S. Stewart, and H. J. Flint, "*Roseburia intestinalis* sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces," *Int. J. Syst. Evol. Microbiol.*, vol. 52, no. 5, pp. 1615–1620, 2002.

[58] C. Chassard and A. Bernalier-Donadille, "H2 and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut," *FEMS Microbiol. Lett.*, vol. 254, no. 1, pp. 116–122, 2006.

[59] A. B. Boraston, "The interaction of carbohydrate-binding modules with insoluble non-crystalline cellulose is enthalpically driven," *Biochem. J.*, vol. 385, no. Pt 2, pp. 479–484, Jan. 2005.

[60] C. Chassard, V. Goumy, M. Leclerc, C. Del'homme, and A. Bernalier-Donadille, "Characterization of the xylan-degrading microbial community from human faeces," *FEMS Microbiol. Ecol.*, vol. 61, no. 1, pp. 121–131, 2007.

[61] C. Mirande *et al.*, "Dietary fibre degradation and fermentation by two xylanolytic bacteria *Bacteroides xylanisolvens* XB1AT and *Roseburia intestinalis* XB6B4 from the human intestine," *J. Appl. Microbiol.*, vol. 109, no. 2, pp. 451–460, 2010.

[62] E. C. M. Leitch, A. W. Walker, S. H. Duncan, G. Holtrop, and H. J. Flint, "Selective colonization of insoluble substrates by human faecal bacteria," *Environ. Microbiol.*, vol. 9, no. 3, pp. 667–679, 2007.

[63] M. L. Leth *et al.*, "Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut," *Nat. Microbiol.*, vol. 3, no. 5, pp. 570–580, 2018.

[64] M. S. Desai *et al.*, "A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility," *Cell*, vol. 167, no. 5, p. 1339–1353.e21, 2016.

[65] K. P. Scott, J. C. Martin, S. H. Duncan, and H. J. Flint, "Prebiotic stimulation of human colonic butyrate-producing bacteria and bifidobacteria, in vitro," *FEMS Microbiol. Ecol.*, vol. 87, no. 1, pp. 30–40, 2014.

[66] S. Rosa *et al.*, "Molecular and Biochemical Basis for Complex β-Mannans Catabolism by the Human Gut Firmicutes Bacterium *Roseburia intestinalis*," *Nat. Commun.*, 2019.

[67] D. W. Cockburn and N. M. Koropatkin, "Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease," *J. Mol. Biol.*, vol. 428, no. 16, pp. 3230–3252, 2016.

[68] A. El Kaoutari, F. Armougom, J. I. Gordon, D. Raoult, and B. Henrissat, "The abundance and variety of carbohydrate-active enzymes in the human gut microbiota.," *Nat. Rev. Microbiol.*, vol. 11, no. 7, pp. 497–504, 2013.

[69] N. I. McNeil, "The contribution of the large intestine to energy supplies in man.," *Am. J. Clin. Nutr.*, vol. 39, no. 2, pp. 338–42, Feb. 1984.

[70] N. Koropatkin, E. Cameron, and E. Martens, "How glycan metabolism shapes the human gut microbiota," *Nat Rev Microbiol*, vol. 10, 2012.

[71] L. A. David *et al.*, "Diet rapidly and reproducibly alters the human gut microbiome," *Nature*, vol. 505, no. 7484, pp. 559–563, 2013.

[72] N. M. Koropatkin, E. a. Cameron, and E. C. Martens, "How glycan metabolism shapes the human gut microbiota," *Nat. Rev. Microbiol.*, vol. 10, no. 5, pp. 323–335, 2012.

[73] A. A. Salyers, S. E. West, J. R. Vercellotti, and T. D. Wilkins, "Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon.," *Appl. Environ. Microbiol.*, vol. 34, no. 5, pp. 529–33, Nov. 1977.

[74] A. A. Salyers, J. R. Vercellotti, S. E. West, and T. D. Wilkins, "Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon.," *Appl. Environ. Microbiol.*, vol. 33, no. 2, pp. 319–22, Feb. 1977.

[75] "Joint FAO/WHO Food Standards Programme, Secretariat of the CODEX Alimentarius Commission: CODEX Alimentarius (CODEX) Guidelines on Nutrition Labeling CAC/GL 2–1985 as Last Amended 2010. Rome: FAO; 2010."

[76] D. Mudgil and S. Barak, "Composition, properties and health benefits of indigestible carbohydrate polymers as dietary fiber: A review," *Int. J. Biol. Macromol.*, vol. 61, pp. 1–6, 2013.

[77] B. R. Hamaker and Y. E. Tuncil, "A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota," *J. Mol. Biol.*, vol. 426, no. 23, pp. 3838–3850, 2014.

[78] M. B. Sticklen, "Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol," *Nat Rev Genet*, vol. 9, no. 6, pp. 433–443, Jun. 2008.

[79] J. Slavin JL, Brauer PM, Marlett, "Neutral detergent fiber, hemicellulose and cellulose digestibility in human subjects," *J. Nutr.*, vol. 111, no. 2, pp. 287–297, 1981.

[80] G. Costa and I. Plazanet, "Plant cell wall , a challenge for its characterisation," *Adv. Biol. Chem.*, vol. 6, no. 3, pp. 70–105, 2016.

[81] H. V Scheller and P. Ulvskov, "Hemicelluloses," *Annu Rev Plant Biol*, vol. 61, 2010.

[82] H. J. Gilbert, "The biochemistry and structural biology of plant cell wall deconstruction.," *Plant Physiol.*, vol. 153, no. 2, pp. 444–455, 2010.

[83] A. Ebringerová and T. Heinze, "Xylan and xylan derivatives - Biopolymers with valuable properties, 1: Naturally occurring xylans structures, isolation procedures and properties,"

Macromol. Rapid Commun., vol. 21, no. 9, pp. 542–556, 2000.

[84] T. Collins, C. Gerday, and G. Feller, "Xylanases, xylanase families and extremophilic xylanases," *FEMS Microbiol. Rev.*, vol. 29, no. 1, pp. 3–23, 2005.

[85] J. Agger, A. Viksø-Nielsen, and A. S. Meyer, "Enzymatic xylose release from pretreated corn bran arabinoxylan: Differential effects of deacetylation and deferuloylation on insoluble and soluble substrate fractions," *J. Agric. Food Chem.*, vol. 58, no. 10, pp. 6141–6148, 2010.

[86] N. R. Swamy, N. Ramakrishnaiah, P. P. Kurien, and P. V Salimath, "Studies on carbohydrates of red gram (*Cajanus cajan*) in relation to milling," *J. Sci. Food Agric.*, vol. 57, no. 3, pp. 379–390, Jan. 1991.

[87] Q. K. Beg, M. Kapoor, L. Mahajan, and G. S. Hoondal, "Microbial xylanases and their industrial applications: A review," *Appl. Microbiol. Biotechnol.*, vol. 56, no. 3–4, pp. 326–338, 2001.

[88] M. J. Peña *et al.*, "Arabidopsis irregular xylem8 and irregular xylem9: Implications for the Complexity of Glucuronoxylan Biosynthesis," *Plant Cell*, vol. 19, no. 2, pp. 549–563, 2007.

[89] D. A. Brummell and R. Schröder, "Xylan metabolism in primary cell walls," *New Zeal. J. For. Sci.*, vol. 39, no. 1, pp. 125–143, 2009.

[90] A. Ebringerová, "Structural diversity and application potential of hemicelluloses," *Macromol. Symp.*, vol. 232, no. 333, pp. 1–12, 2006.

[91] R. P. de Vries, J. Visser, P. Ronald, R. de Vries, and P., "*Aspergillus* Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides," *Microbiol. Mol. Biol. Rev.*, vol. 65, no. 4, pp. 497–522, 2001.

[92] H. R. Sørensen, S. Pedersen, and A. S. Meyer, "Characterization of solubilized arabinoxylo-oligosaccharides by MALDI-TOF MS analysis to unravel and direct enzyme catalyzed hydrolysis of insoluble wheat arabinoxylan," *Enzyme Microb. Technol.*, vol. 41, no. 1–2, pp. 103–110, Jul. 2007.

[93] R. R. Schendel, M. R. Meyer, and M. Bunzel, "Quantitative Profiling of Feruloylated Arabinoxylan Side-Chains from Graminaceous Cell Walls.," *Front. Plant Sci.*, vol. 6, no. January, p. 1249, 2015.

[94] A. Rogowski *et al.*, "Glycan complexity dictates microbial resource allocation in the large intestine.," *Nat. Commun.*, vol. 6, no. May, p. 7481, 2015.

[95] P. S. Chauhan, N. Puri, P. Sharma, and N. Gupta, "Mannanases: Microbial sources, production, properties and potential biotechnological applications," *Appl. Microbiol. Biotechnol.*, vol. 93, no. 5, pp. 1817–1830, 2012.

[96] S. Dhawan and J. Kaur, "Microbial mannanases: an overview of production and applications," *Crit. Rev. Biotechnol.*, vol. 27, no. 4, pp. 197–216, 2007.

[97] L. R. S. Moreira and E. X. F. Filho, "An overview of mannan structure and mannan-degrading enzyme systems," *Appl. Microbiol. Biotechnol.*, vol. 79, no. 2, pp. 165–178, 2008.

[98] V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat, "The carbohydrate-active enzymes database (CAZy) in 2013," *Nucleic Acids Res.*, vol. 42, no. D1, pp. 490–495, 2014.

[99] B. I. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics," *Nucleic Acids Res.*, vol. 37, no. SUPPL. 1, pp. 233–238, 2009.

[100] G. J. Davies and B. Henrissat, "Structures and Mechanisms of Glycosyl Hydrolases," *Structure*, vol. 3, no. 9, pp. 853–859, 1995.

[101] B. Henrissat and A. Bairoch, "Updating the sequence-based classification of glycosyl hydrolases," *Biochem. J.*, vol. 316, no. 2, pp. 695–696, 1996.

[102] D. E. Koshland, "Stereochemistry and the mechanism of enzymatic reactions," *Biol. Rev.*, vol. 28, no. 4, pp. 416–436, Nov. 1953.

[103] D. W. Abbott and A. B. Boraston, *Quantitative approaches to the analysis of carbohydrate-binding module function*, 1st ed., vol. 510. Elsevier Inc., 2012.

[104] A. B. Boraston, D. N. Bolam, H. J. Gilbert, and G. J. Davies, "Carbohydrate-binding modules: fine-tuning polysaccharide recognition," *Biochem. J.*, vol. 382, no. 3, pp. 769–781, 2004.

[105] S. Armenta, S. Moreno-Mendieta, Z. Sánchez-Cuapio, S. Sánchez, and R. Rodríguez-Sanoja, "Advances in molecular engineering of carbohydrate-binding modules," *Proteins Struct. Funct. Bioinforma.*, vol. 85, no. 9, pp. 1602–1617, Sep. 2017.

[106] N. Din, H. G. Damude, N. R. Gilkes, R. C. Miller, R. a Warren, and D. G. Kilburn, "C1-Cx revisited: intramolecular synergism in a cellulase.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 91, no. 24, pp. 11383–11387, 1994.

[107] H. J. Gilbert, J. P. Knox, and A. B. Boraston, "Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules," *Curr Opin Struct Biol*, vol. 23, 2013.

[108] B. M. Campos *et al.*, "A Novel Carbohydrate-binding Module from Sugar Cane Soil Metagenome Featuring Unique Structural and Carbohydrate Affinity Properties," *J. Biol. Chem.* , vol. 291, no. 45, pp. 23734–23743, Nov. 2016.

[109] N. Georgelis, N. H. Yennawar, and D. J. Cosgrove, "Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin," *Proc. Natl. Acad. Sci.*, vol. 109, no. 37, p. 14830 LP-14835, Sep. 2012.

[110] A. B. Boraston, M. Healey, J. Klassen, E. Ficko-Blean, A. Lammerts van Bueren, and V. Law, "A structural and functional analysis of alpha-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition.," *J. Biol. Chem.*, vol. 281, no. 1, pp. 587–98, Jan. 2006.

[111] D. N. Bolam *et al.*, "Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* Xylanase 11A," *Biochemistry*, vol. 40, no. 8, pp. 2468–2477, 2001.

[112] A. M. Nakamura, A. S. Nascimento, and I. Polikarpov, "Structural diversity of carbohydrate esterases," *Biotechnol. Res. Innov.*, pp. 1–17, 2017.

[113] D. Dodd and I. K. Cann, "Enzymatic deconstruction of xylan for biofuel production," *Glob Chang. Biol Bioenergy*, vol. 1, no. 1, pp. 2–17, 2009.

[114] P. Biely, S. Singh, and V. Puchart, "Towards enzymatic breakdown of complex plant xylan structures: State of the art," *Biotechnol. Adv.*, vol. 34, no. 7, pp. 1260–1274, 2016.

[115] F. J. St John, J. C. Hurlbert, J. D. Rice, J. F. Preston, and E. Pozharski, "Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase," *J. Mol. Biol.*, vol. 407, no. 1, pp. 92–109, Mar. 2011.

[116] F. I. J. Pastor, Ó. Gallardo, J. Sanz-apaicio, and P. Díaz, *Chapter 5, Xylanases: Molecular properties and applications*. 2007.

[117] Y. Honda and M. Kitaoka, "A family 8 glycoside hydrolase from *Bacillus halodurans* C-125 (BH2105) is a reducing end xylose-releasing exo-oligoxylanase," *J. Biol. Chem.*, vol. 279, no. 53, pp. 55097–55103, 2004.

[118] S. V Valenzuela, S. Lopez, P. Biely, J. Sanz-aparicio, and F. I. J. Pastor, "Active on Branched Xylooligosaccharides," vol. 82, no. 17, pp. 5116–5124, 2016.

[119] D. Nurizzo, T. Nagy, H. J. Gilbert, and G. J. Davies, "The Structural Basis for Catalysis and Specificity of the Pseudomonas cellulosa α-Glucuronidase, GlcA67A," *Structure*, vol. 10, no. 4, pp. 547–556, Apr. 2002.

[120] K. Kolenová, O. Ryabova, M. Vršanská, and P. Biely, "Inverting character of family GH115 a-glucuronidases," *FEBS Lett.*, vol. 584, no. 18, pp. 4063–4068, 2010.

[121] K. M. J. Van Laere, C. H. L. Voragen, T. Kroef, L. A. M. den Broek, G. Beldman, and A. G. J. Voragen, "Purification and mode of action of two different arabinoxylan arabinofuranohydrolases from Bifidobacterium adolescentis DSM 20083," *Appl. Microbiol.*

Biotechnol., vol. 51, no. 5, pp. 606–613, 1999.

[122] K. M. J. Van Laere, G. Beldman, and A. G. J. Voragen, "A new arabinofuranohydrolase from Bifidobacterium adolescentis able to remove arabinosyl residues from double-substituted xylose units in arabinoxylan," *Appl. Microbiol. Biotechnol.*, vol. 47, no. 3, pp. 231–235, Mar. 1997.

[123] H. Ferré, A. Broberg, J. Ø. Duus, and K. K. Thomasen, "A novel type of arabinoxylan arabinofuranohydrolase isolated fromgerminated barley," *Eur. J. Biochem.*, vol. 38, no. 2, pp. 189–198, 2000.

[124] P. Biely, M. Mastihubová, and V. Puchart, "The vicinal hydroxyl group is prerequisite for metal activation of *Clostridium thermocellum* acetylxylan esterase," *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1770, no. 4, pp. 565–570, Apr. 2007.

[125] F. A. Adesioye, T. P. Makhalanyane, P. Biely, and D. A. Cowan, "Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases," *Enzyme Microb. Technol.*, vol. 93–94, pp. 79–91, 2016.

[126] S. Malgas, J. S. van Dyk, and B. I. Pletschke, "A review of the enzymatic hydrolysis of mannans and synergistic interactions between β-mannanase, β-mannosidase and α-galactosidase," *World J. Microbiol. Biotechnol.*, vol. 31, no. 8, pp. 1167–1175, 2015.

[127] L. E. Tailford *et al.*, "Understanding how diverse β-mannanases recognize heterogeneous substrates," *Biochemistry*, vol. 48, no. 29, pp. 7009–7018, 2009.

[128] P. Ademark *et al.*, "Hydrolytic properties of a beta-mannosidase purified from *Aspergillus niger*.," *J. Biotechnol.*, vol. 75, no. 2–3, pp. 281–289, 1999.

[129] H. J. Gilbert, H. Stålbrand, and H. Brumer, "How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation," *Curr. Opin. Plant Biol.*, vol. 11, no. 3, pp. 338–348, 2008.

[130] M. Tenkanen, J. Thornton, and L. Viikaria, "An acetylglucomannan esterase of Aspergillus oryzae; purification, characterization and role in the hydrolysis of O-acetyl-galactoglucomannan," *J. Biotechnol.*, vol. 42, no. 3, pp. 197–206, Oct. 1995.

[131] M. Ejby, F. Fredslund, A. Vujicic-Zagar, B. Svensson, D. J. Slotboom, and H. M. Abou, "Structural basis for arabinoxylo-oligosaccharide capture by the probiotic Bifidobacterium animalis," *Mol Microbiol*, vol. 90, pp. 1100–1112, 2013.

[132] P. M. Jones and A. M. George, "The ABC transporter structure and mechanism: Perspectives on recent research," *Cell. Mol. Life Sci.*, vol. 61, no. 6, pp. 682–699, 2004.

[133] A. L. Davidson, E. Dassa, C. Orelle, and J. Chen, "Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems," *Microbiol. Mol. Biol. Rev.*, vol. 72, no. 2, pp. 317–364, 2008.

[134] T. Eitinger, D. A. Rodionov, M. Grote, and E. Schneider, "Canonical and ECF-type ATP-binding cassette importers in prokaryotes: Diversity in modular organization and cellular functions," *FEMS Microbiol. Rev.*, vol. 35, no. 1, pp. 3–67, 2011.

[135] D. C. Rees, E. Johnson, and O. Lewinson, "ABC transporters: the power to change," *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 3, pp. 218–227, 2009.

[136] J. ter Beek, A. Guskov, and D. J. Slotboom, "Structural diversity of ABC transporters," *J. Gen. Physiol.*, vol. 143, no. 4, pp. 419–435, 2014.

[137] R. P. A. Berntsson, S. H. J. Smits, L. Schmitt, D. J. Slotboom, and B. Poolman, "A structural classification of substrate-binding proteins," *FEBS Lett.*, vol. 584, no. 12, pp. 2606–2617, 2010.

[138] B. Mao, M. R. Pear, A. J. Mccammon, and F. Quiocho, "Hinge-bending in L-arabinose-binding protein . The " Venus's-flytrap " model," *Chem. Biol.*, vol. 257, no. February 10, pp. 1131–1133, 1982.

[139] M. L. Oldham and J. Chen, "Crystal Structure of the Maltose Transporter in a Pretranslocation Intermediate State," *Science (80-.).*, vol. 332, no. 6034, pp. 1202–1205, 2011.

[140] J. Xu *et al.*, "Evolution of Symbiotic Bacteria in the Distal Human Intestine," *PLOS Biol.*, vol.

5, no. 7, p. e156, Jun. 2007.

[141] M. H. Foley, D. W. Cockburn, and N. M. Koropatkin, "The Sus operon: a model system for starch uptake by the human gut Bacteroidetes," *Cell. Mol. Life Sci.*, vol. 73, no. 14, pp. 2603–2617, Jul. 2016.

[142] E. C. Martens, N. M. Koropatkin, T. J. Smith, and J. I. Gordon, "Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm," *J. Biol. Chem.*, vol. 284, no. 37, pp. 24673–24677, Sep. 2009.

[143] E. Tancula, M. J. Feldhaus, L. A. Bedzyk, and A. A. Salyers, "Location and characterization of genes involved in binding of starch to the surface of Bacteroides thetaiotaomicron," *J. Bacteriol.*, vol. 174, no. 17, pp. 5609–5616, Sep. 1992.

[144] J. M. Grondin, K. Tamura, G. Déjean, D. W. Abbott, and H. Brumer, "Polysaccharide Utilization Loci: Fueling Microbial Communities," *J. Bacteriol.*, vol. 199, no. 15, pp. e00860-16, Aug. 2017.

[145] A. J. Glenwright *et al.*, "Structural basis for nutrient acquisition by dominant members of the human gut microbiota," *Nature*, vol. 541, p. 407, Jan. 2017.

[146] A. S. Tauzin *et al.*, "Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont," *MBio*, vol. 7, no. 2, pp. e02134-15, May 2016.

[147] N. M. Koropatkin, E. C. Martens, J. I. Gordon, and T. J. Smith, "Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices," *Structure*, vol. 16, no. 7, pp. 1105–1115, Jul. 2008.

[148] G. R. Gibson *et al.*, "Expert consensus document: The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics," *Nat. Rev. Gastroenterol. Hepatol.*, vol. 14, no. 8, pp. 491–502, 2017.

[149] G. R. Gibson and M. B. Roberfroid, "Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics.," *J. Nutr.*, vol. 125, no. 6, pp. 1401–1412, 1995.

[150] R. A. Rastall and G. R. Gibson, "Recent developments in prebiotics to selectively impact beneficial microbes and promote intestinal health," *Curr. Opin. Biotechnol.*, vol. 32, pp. 42–46, 2015.

[151] S. Perrin, C. Fougnies, J. P. Grill, H. Jacobs, and F. Schneider, "Fermentation of chicory fructo-oligosaccharides in mixtures of different degrees of polymerization by three strains of bifidobacteria," *can j Microbiol*, vol. 763, pp. 759–763, 2002.

[152] J. I. Sanchez *et al.*, "Arabinoxylan-oligosaccharides (AXOS) affect the protein/carbohydrate fermentation balance and microbial population dynamics of the Simulator of Human Intestinal Microbial Ecosystem," *Microb. Biotechnol.*, vol. 2, no. 1, pp. 101–113, 2009.

[153] Y. J. Goh and T. R. Klaenhammer, "Genetic Mechanisms of Prebiotic Oligosaccharide Metabolism in Probiotic Microbes," *Annu. Rev. Food Sci. Technol.*, vol. 6, no. 1, pp. 137–156, 2015.

[154] H. J. Flint, S. H. Duncan, K. P. Scott, and P. Louis, "Links between diet, gut microbiota composition and gut metabolism," *Proc. Nutr. Soc.*, vol. 74, no. 01, pp. 13–22, 2015.

[155] J. A. Krumbeck, M. X. Maldonado-Gomez, A. E. Ramer-Tait, and R. W. Hutkins, "Prebiotics and synbiotics," *Curr. Opin. Gastroenterol.*, vol. 32, no. 2, pp. 110–119, 2016.

[156] R. A. Rastall, "Functional Oligosaccharides: Application and Manufacture," *Annu. Rev. Food Sci. Technol.*, vol. 1, no. 1, pp. 305–339, 2010.

[157] D. W. Cockburn *et al.*, "Molecular details of a starch utilization pathway in the human gut symbiont Eubacterium rectale," *Mol. Microbiol.*, vol. 95, no. 2, pp. 209–230, Jan. 2015.

[158] A. G. Ramsay, K. P. Scott, J. C. Martin, M. T. Rincon, and H. J. Flint, "Cell-associated α-amylases of butyrate-producing Firmicute bacteria from the human colon," *Microbiology*, vol. 152, no. 11, pp. 3281–3290, 2006.

[159] X. Ze *et al.*, "Unique organization of extracellular amylases into amylosomes in the resistant

starch-utilizing human colonic firmicutes bacterium *Ruminococcus bromii*," *MBio*, vol. 6, no. 5, pp. 1–11, 2015.

[160] K. De Paepe, F.-M. Kerckhof, J. Verspreet, C. M. Courtin, and T. Van de Wiele, "Inter-individual differences determine the outcome of wheat bran colonization by the human gut microbiome," *Environ. Microbiol.*, vol. 00, pp. 1–17, 2017.

[161] N. A. Treptow and H. A. Shuman, "Genetic evidence for substrate and periplasmic-binding-protein recognition by the MalF and MalG proteins, cytoplasmic membrane components of the Escherichia coli maltose transport system.," *J. Bacteriol.*, vol. 163, no. 2, p. 654 LP-660, Aug. 1985.

[162] K. J. O´Connell *et al.*, "Metabolism of Four α-Glycosidic Linkage-Containing Oligosaccharides by Bifidobacterium breve UCC2003," *Appl. Environ. Microbiol.*, vol. 79, no. 20, p. 6280 LP-6292, Oct. 2013.

[163] F. Cuskin *et al.*, "Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism," *Nature*, vol. 517, p. 165, Jan. 2015.

[164] K. D'hoe *et al.*, "Integrated culturing, modeling and transcriptomics uncovers complex interactions and emergent behavior in a three-species synthetic gut community," *Elife*, vol. 7, Oct. 2018.

[165] M. A. Mahowald *et al.*, "Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla," *Proc Natl Acad Sci U S A*, vol. 106, no. 14, pp. 5859–5864, 2009.

[166] Y. E. Tuncil, Y. Xiao, N. T. Porter, B. L. Reuhs, E. C. Martens, and B. R. Hamaker, "Reciprocal Prioritization to Dietary Glycans by Gut Bacteria in a Competitive Environment Promotes Stable Coexistence," *MBio*, vol. 8, no. 5, pp. e01068-17, Oct. 2017.

[167] N. A. Pudlo, K. Urs, S. S. Kumar, J. B. German, D. A. Mills, and E. C. Martens, "Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans," *MBio*, vol. 6, no. 6, pp. e01282-15, Dec. 2015.

[168] S. J. Reid and V. R. Abratt, "Sucrose utilisation in bacteria: genetic organisation and regulation," *Appl. Microbiol. Biotechnol.*, vol. 67, no. 3, pp. 312–321, 2005.

[169] L. Swint-Kruse and K. S. Matthews, "Allostery in the LacI/GalR family: variations on a theme," *Curr. Opin. Microbiol.*, vol. 12, no. 2, pp. 129–137, Apr. 2009.

[170] E. Madland *et al.*, "1H, 13C and 15N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from *Roseburia intestinalis*," *Biomol. NMR Assign.*, Sep. 2018.

Appendix 1

```
# Program: needle
# Aligned_sequences: 2
# 1: EEU99941.1 ROSINTL182_08194
# 2: EEV02614.1 ROSINTL182_05471
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
# Length: 614
# Identity:     114/614 (18.6%)
# Similarity:   181/614 (29.5%)
# Gaps:         289/614 (47.1%)
# Score: 371.5
#=======================================

EEU99941.1         1 MSGPVAPKDPEKDRSYFYIMKEKETFGSLQTQGEYQGRGVQFIYESDGRL     50

EEV02614.1         1 --------------------------------------------------       0

EEU99941.1        51 ESSAEVTGEVCDEEILKKLGTVEGFKSLVHSIGISVEMEHSREPVTFVFQ    100

EEV02614.1         1 --------------------------------------------------       0

EEU99941.1       101 MYGKEDLYGGGTLIETELRGDGAEVRITLDTVKWKTDDDVPGQIRFVFET    150

EEV02614.1         1 --------------------------------------------------       0

EEU99941.1       151 PEQSARVNVRFFLKDGFFVPKPQEERVVDMESHGYQEMIERSLLSMGDAG    200
                                                ||   ||...|..:.:.|...
EEV02614.1         1 ------------------------------ME---YQIKYENGIANRGCLY     18

EEU99941.1       201 RIRRVVEKARAGEPVTIAYIGGSITQGA-GAVPLHTQCYA---YRFWKAF    246
                     |:::|:::|:|||.:.||::|||||||: .:.|  ..|||   |.:||
EEV02614.1        19 RLKKVMDRAKAGEALNIAFLGGSITQGSLSSKP--ELCYAYHVYEWWK--     64

EEU99941.1       247 AGKYGKNNNVKLIKAGVGGTPSELGMIRFERDVLRDGKEKPDLVVVEFAV    296
                       |.....:...|.||:|||.|:.|:.|.|.|:|   .::||.|::||:|
EEV02614.1        65 --KTFPQADFTYINAGIGGTTSQFGVARAEADLL---SKEPDFVIIEFSV    109

EEU99941.1       297 NDEGDETKGRCYESLVTKILSMPDAPAVLLLFAVFANDW-NLQERLAPVG    345
                     ||:..|.....||.||.|:.:....|||||:..||.|:. |.|.....:.
EEV02614.1       110 NDDSTEHFMETYEGLVRKVYTSKTKPAVLLVHNVFYNNGANAQLMHGRIA    159

EEU99941.1       346 ERYQLPMVSIRDAVTPQFRQTKDRVVSKNQFFYDAFHPTNLGHKIMADCL    395
                     ..|.||.||::..:.|:....:   :...:...|..||.:.||.::|..:
EEV02614.1       160 RHYNLPAVSMQSTIYPEVVAGR---IENREITPDDLHPNDAGHALVASVI    206

EEU99941.1       396 MYLIDRA-------VCEPD----ILRRMHEKPVYGDEFAQVKLLDRRDGY    434
                     .|.:|:.       ..|||    :.:..:||.:            |....
EEV02614.1       207 TYFLDKVKTEDATEQSEPDYPAPLTKNTYEKSI------------RHQNS    244

EEU99941.1       435 ERAKICCGAFSGTDQELQCVEMDDSLTPVPEFPYNWQYDGANGRSEDAFQ    484
                     :...:|.|..:.|..:....:.         |.:.|    ...:..|:..
EEV02614.1       245 DENVVCHGFVADTSAQRDITDC---------FKHGW----TASKKGDSIT    281

EEU99941.1       485 MDI--------YCRSLLL------IFKDSGAVDAGRADVFVDDT--KVLT    518
                     :|:        |.:|:.|      |..|..|..|.|.|...|:|  ..|.
EEV02614.1       282 LDVEGCNISVQYRKSVKLPAPVAEIIVDGDAEHAVRLDANFDETWGDKLE    331

EEU99941.1       519 ADPHVNGWTHCNPVILLEEKESGWHQVRIQMTP---------------GE    553
                     .|            .:||..|:..|:|.:::|.               |.
EEV02614.1       332 LD------------TILEHGENKVHKVEVRLTETHENDAVPFYLVSVIGS    369

EEU99941.1       554 EEKKFTILGFGYVE    567
                     .||
EEV02614.1       370 SEK-----------    37
```